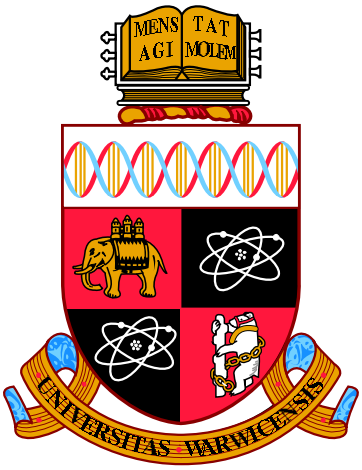


Visual Attention for High-Fidelity Imaging

Timothy Bradley
MEng (Hons)

A thesis submitted to the University of Warwick
for the degree of
Doctor of Philosophy in Engineering

WMG
University of Warwick
September 2017

Contents

List of Figures

List of Tables

Acknowledgements

I'm writing this at the last possible moment, as is my way, and I'm sure I'm going to miss some people out. What follows is a (non-exhaustive) list of people who carried me kicking and screaming across the finish line.

Firstly I would like to thank my supervisors Alan Chalmers and Kurt Debattista. Alan for always being there when I needed to talk about my imminent failure or the absolute certainty that everyone would realise how terrible my work was. Kurt for sitting and working through my ideas with me, and for teaching me statistics through osmosis.

Tom Bashford-Rogers, I think most of the ideas for my work came out of our late night chats when we were the only people left in the office. Also thank you for saving me a lot of time reading papers by saying "It's already been done," more times than I can remember. Carlo Harvey, thank you for always being kind when I was being dumb, your excellent 3D models, and for sitting with me debugging my code for hours.

I have to thank my friends Amar, Jon, and Josh. We started our degrees on the same day, and submitted within five minutes of each other. We've been through so much together that I can't really cover it here so I'll say: Jon, thanks for playing Rocket League with me when we should have been working; Josh, thank you for losing our prank war; Amar, thank you for building a PC for me.

I also want to than everyone in the Visualisation group for their friendship and for giving me endless amounts of proofreading to do, your examples helped me write this thesis. I'd also like to thank everyone who volunteered to take part in my experiments and made my research possible.

To my friends outside university, thank you for providing a break from the stress. I

can't mention everyone but there are a few I feel the need to mention by name. Tom, thank you for keeping me sane with our incredibly unhealthy weekend gaming sessions. Rhys, for being one of my closest friends throughout this. Chris, for being there anytime I needed or wanted to talk, I feel like we've known each other a lot longer than we have.

I couldn't have done this without my family. My parents, John and Isabel Bradley, for encouraging me to study engineering and always supporting me through any questionable life choice I made. Also for always being there when I didn't know what to do. Also to my brothers, Bob and Nick for teaching me how to be awesome.

Finally Mihika, nothing I can say will ever be enough. Thanks for picking up the pieces every time I fell apart, I love you bubba.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

Publications

The work in this thesis has contributed to the following publication:

- Tim Bradley, Kurt Debattista, Thomas Bashford-Rogers, Carlo Harvey, E Doukakis, and Alan Chalmers. Selective brdfs for high fidelity rendering. In *Proceedings of the conferece on Computer Graphics & Visual Computing*, pages 57–64. Eurographics Association, 2016

Abstract

Models of visual attention have many applications including but not limited to rendering, advertising, graphic design and road safety. The rise of high fidelity imaging technologies, such as high dynamic range content and physically-based rendering have created a need for more targeted models, however the data necessary for their creation is sparse.

This thesis aims to expand the applicability of visual attention frameworks for high-fidelity imaging both by introducing a new selective rendering method for adaptively adjusting the quality of rendered scenes and developing the necessary tools to validate existing and future models in high fidelity domains.

This thesis first presents a method for exploiting visual attention, in a Physically-Based-Rendering (PBR) pipeline, by adjusting complexity of Bidirectional Reflectance Distribution Functions (BRDFs) in unimportant image regions. Thus, the presented method substitutes high accuracy, high cost models with low accuracy, low cost models in less salient regions. The efficacy of this method is evaluated through a subjective rating experiment. The results of the psychophysical experiment found some significant confusion between the hybrid and references images, which suggests that this can be employed as a tool to reduce computational costs.

Furthermore, this thesis presents an experiment to assess the effect of high luminance levels on the viewing strategies of observers. This is accomplished through the creation of an HDR eye-tracking dataset consisting of eighty HDR images, shown at four distinct brightness levels. A statistical analysis of the resulting fixation density maps found that the reliability of LDR eye tracking data decreases as the peak brightness of an images increases. This suggests the need for reliable HDR eye-tracking datasets.

Finally, this thesis presents an eye-tracking experiment and subjective survey to analyse the interaction of ambient light levels and screen brightness on visual fatigue and visual saliency. Results of the experiment show an increase in similarity between HDR and LDR fixations as environmental illumination increases, this is of particular note as standard practice calls for eye-tracking dataset to be captured in dark environments.

Chapter 1

Introduction

The goal of high-fidelity imaging is to precisely represent, recreate, or capture the real world, or to present a virtual scene indistinguishable from reality. High-fidelity imaging encompasses the fields of rendering, stereo imaging, virtual reality, volumetric displays, and High Dynamic Range (HDR) imaging.

These fields have a requirement for the display of realistic images whether for a more precise representation of real scenes, for photography and cinematography, or for capturing more reliable data for scientific purposes, such as image-based lighting, or for the rendering of realistic virtual scenes used in prototyping, architecture, and video games. Figure 1.1 shows an example of a high-fidelity rendered scene.

A factor common to these areas of high-fidelity imaging, is the burden of processing and storing a large amount of data. However, much of this processing may not be necessary when computing images for human consumption, because the bandwidth of the Human Visual System (HVS) is inherently limited. How a human attends to a scene and thus what a human actually observes has been the subject of great interest to psychologists, computer scientists and neuroscientists [FRC10]. Beyond academic circles, attracting or avoiding the focus of visual attention has a wide range of commercial, industrial, and even military uses, including but not limited to advertising, web-design, road signage, film, camouflage, gaming, user interfaces, and rendering. Models of visual attention attempt to predict this process either algorithmically through assumptions derived from a neurological basis [IKN98], or via a trained machine learning model

Figure 1.1: A rendered image of a lounge scene

[BDDPN15]. Algorithmic models, such as the original implementation of Itti and Koch [IKN98], apply filters to images targeting known preattentive features [HE12], identified from neurological research and empirical testing. The features utilised by Koch and Ullman [KU87] are colour, orientation, and local luminance. This luminance feature is primarily centred on detecting local contrast, much like in feature detectors such as Lowe's Scale-Invariant Feature Transform (SIFT) detector [Low04], as sharp changes in local brightness often represent the edges of objects. The emergence of high-fidelity imaging has brought with it an increased number of potential preattentive features: such as depth, in the case of stereo-imaging; reflection artefacts in rendering; and the brightness of the screen and environment in HDR imaging.

Visual attention models take in an image and output a saliency map, see Figure 1.2, which is a two-dimensional probability mass function stored as an image wherein pixel values represent the probability that the respective pixel in the original image will be attended. In the case of temporal saliency models, videos are taken as inputs and a number of saliency maps are generated. These maps are used to direct content placement in areas such as photography, film and advertising, and to direct resources in

rendering and gaming. The accuracy of attention prediction is therefore the key measure of performance for visual attention models, particularly in rendering, as resources incorrectly distributed in the quest of reducing computational burdens are wasted.

Figure 1.2: A rendered conference scene (left) and its corresponding saliency map (right)

A significant amount of data collection is required for adequate targeting, validation, and training of visual attention models, but while many datasets exist for standard imaging [WS13], few exist for high-fidelity imaging and targeted environments. Targeted environments are those with known ambient conditions, such as a dark cinema theatre, or a virtual reality head mounted device, for which models can be specifically tailored. Furthermore, the effects of such environments and content on visual attention have not been adequately studied. The datasets collected for standard imagery suffers from this, as data is collected in ideal laboratory conditions. This is a significant oversight, as much like high-fidelity imaging, the value of visual attention models are in their accuracy and failing to account for a significant attentive feature can only serve to make a model less accurate.

1.1 Research Problem

Visual attention models, whether designed or trained, are feature-driven, that is they rely not only on their ability to discriminate between important regions and random noise but also on the correct choice of features for which to search. Visual attention, thus far, has been insufficiently adapted for use in high-fidelity imaging, with features specific to the various new display technologies and environments under-examined. The primary aim of this thesis is to investigate the effects of various high-fidelity imaging

techniques on visual attention, providing a baseline analysis of features common to these techniques from which future work can build.

1.1.1 Research Question

This thesis aims to expand the empirical foundation of visual attention for high-fidelity imaging methods, by providing new methods for efficient generation of virtual scenes, and refining the empirical basis for targeted models of visual attention for high-fidelity content.
The research question can thus be stated as follows:

- How can in-depth knowledge of visual attention be used effectively when generating high-fidelity imagery?

To answer this there are three objectives:

- Develop and test a selective rendering method which utilises Bidirectional Reflectance Distribution Functions (BRDFs).
- Create an HDR eye tracking data set with multiple uniform enforced brightnesses.
- Analyse the effect of HDR and ambient light on visual attention and visual fatigue.

1.1.2 Overview

To tackle the research question three major aspects follow these objectives:

The study and availability of high-fidelity imaging has, with the advent of stereo displays, high dynamic range imaging, virtual reality, and physically-based rendering become increasingly relevant in recent years, and remains a growing field. Rendering, while a computationally expensive process, is of fundamental importance in the video game, architecture, product design, and film industries. Several strategies have been developed to mitigate a portion of this cost by exploiting inattentional blindness, such as adjusting polygon counts, coherence algorithms, sampling strategies, and global illumination. As efforts in rendering have moved towards greater realism, the simulation of reflection has become more precise. However as additional effects, such as

self shadowing and shading, are integrated into existing reflectance functions, so the computational cost incurred by these functions has also increased. The method proposed and tested as part of this thesis, in Chapter 5, aims to add to existing techniques of alleviating this computational burden by exploring the perceptual costs of simplifying complex reflection models in unimportant regions.

The field of HDR imaging, has gained significant traction in recent years, with limited HDR functionality being built into modern smartphones, such as the iPhone SE, and games consoles, such as the PS4 Pro. HDR imaging encompasses the capture, storage and display of image and video content containing luminance values beyond what standard, Low Dynamic Lange (LDR) imaging is capable of processing. HDR imaging aims to provide a more faithful reproduction of real scenes, a more accurate data source for other high-fidelity imaging applications such as image based lighting, and an overall more visually pleasing experience. Despite the increasing adoption of this technology, only a small number of visual attention models have been developed for HDR content, with even fewer eye tracking datasets available to study the underlying effects of the high luminance levels produced by HDR displays.

The range of luminance in which the human visual system is capable of perceiving detail concurrently is limited to luminance ratios of $1 : 10^4 \text{cd/m}^2$ [BADC17], without adaptation. Thus with HDR displays capable of displaying luminance of 4,000 and 8,000 cd/m^2 it is imperative to account for the effects of the both high display luminance and ambient illuminance on visual attention as it relates to LDR and HDR content. Traditionally eye tracking data collection and perceptual tests for high-fidelity content are captured in dark environments. Therefore the effects of this are so far unaccounted for, with both high-fidelity and traditional content.

1.2 Outline

- **Chapters 2 and 3** provide an overview of the relevant topics and methods, necessary to give full context to the work described in this thesis.

- **Chapter 4** describes the methodology employed in later chapters to address the

research question and the stated objectives.

- **Chapter 5** describes a method for selectively targeting high quality reflectance functions to areas of high visual importance, and documents the methods and results of a psychophysical experiment analysing its effectiveness.

- **Chapter 6** details an eye tracking study and statistical analysis on the relation of display brightness to the viewing strategies of observers.

- **Chapter 7** documents an eye tracking data study, examining the effects of ambient light and high dynamic range displays on visual attention and visual fatigue.

- Finally, **Chapter 8** presents concluding remarks, outlines potential areas for future work, and discusses the limitations of the work presented in this thesis.

Chapter 2

High Fidelity Imaging

2.1 Introduction

This chapter details the background theory and literature relevant to high fidelity imaging, selective rendering.

2.2 High Dynamic Range Imaging

The Human Visual System (HVS) is capable of observing a much wider range of luminance values simultaneously than a conventional Low Dynamic Range (LDR) camera can capture in a single exposure [RHD+10]. In recent years High Dynamic Range (HDR) imaging has closed the gap and exceeded the eyes' ability to capture scenes of extreme lighting.

Among other things this has applications in feature matching [CARR14], display [Sel13a], and creating dense depth maps [TKS06]. Typical LDR cameras capture very limited illumination levels and parts of the image become over-saturated and under-saturated, removing detail from the image. Table 2.1 provides a list of luminance levels of common environments. The primary goal of HDR imaging is to capture as much information about the scene as possible to aid in a variety of applications from image based lighting to astronomy [MYS95].

Conditions	Illumination (in cd/m^2)
Starlight	10^{-3}
Moonlight	10^{-1}
Indoor lighting	10^{2}
Sunlight	10^{5}
Max. intensity of Standard Computer Displays	10^{2}

Table 2.1: Ambient luminance levels for some common lighting environments [RHD+10][Wan95]

2.2.1 Terminology

This section summarises some HDR terminology relevant to later sections.

Radiance is a measure of the amount of light arriving at a specific area, in a given time, from a given angle. This a particularly useful unit as it describes the amount of light energy that a sensor element or pixel captures.

Luminance is also a measure of light per unit time incident upon a specific area from a particular direction or solid angle but is corrected to correspond to the human visual system. It is measured in candelas per metre squared, cd/m^2.

In photography *Dynamic Range* is the difference between the lightest and darkest regions of an image. It is similar to the contrast ratio of computer displays and televisions, see equation 2.1. However most contrast ratios given in the technical specifications for displays mean little as there is no standard for how to measure the minimum luminance. See [Mor11] for more detail. For example one manufacturer might take the minimum luminance from a black screen with the backlight at its minimum brightness setting and the maximum luminance from the highest backlight setting. The contrast ratio calculated from these two readings would be meaningless as it wouldn't represent what can be seen simultaneously.

$$ContrastRatio(CR) = \frac{Luminance_{max}}{Luminance_{min}} : 1 \tag{2.1}$$

In digital imaging dynamic range is a measure of the difference between the maximum capacity, MAX_f, of a pixel on the camera sensor and the base noise output or read out voltage of the pixel. Peak Signal to Noise Ratio or PSNR is a good way of accounting for this noise. It is a unitless logarithmic measure of the maximum value versus the noise of a system and is therefore measured in decibels, see equation 2.2 [psn13].

$$PSNR = 20log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \tag{2.2}$$

Where MSE is Mean Square Error and can be calculated using equation 2.3.

$$MSE = \frac{1}{mn}\sum_{0}^{m-1}\sum_{0}^{n-1}\|f(i,j) - g(i,j)\|^2 \tag{2.3}$$

Where MAX_f is the maximum pixel value in the original image, f is the original image matrix, g is the matrix of the degraded image. m represents the number of rows of pixels and i represents the index of that row. n represents the number of columns of pixels and j represents the index of that column.

Photographers use f-stops, or f-numbers, as a measure differences in luminance. f-stops describe the ratio of the cameras focal length, f, to the diameter of it's aperture, D:

$$\text{F-Number} = \frac{f}{D} \tag{2.4}$$

The human visual system's response to luminance is non-linear and can be approximated to a square relationship. An increase of one stop implies a doubling in luminance. See equation 2.5 for a mathematical illustration of this.

$$n \text{ stops} = 2^n \text{ Dynamic Range} \tag{2.5}$$

The *bit depth* of an image is the number of bits assigned to each pixel channel or it can mean the number of bits assigned to each pixel. For simplicity from this point on bit depth will refer to the number of bits per channel and *bpp* will be used for bits per pixel. In LDR image formats each channel is represented by an 8 bit integer, this means that each channel can only have 256 possible values ranging from 0-255. Bit depth isn't

connected to dynamic range, 0 and 255 could represent any values. However trying to represent a high dynamic range with 8 bit integers would cause extreme banding and posterisation.

This limits the possible dynamic range of a linearly encoded LDR image to 8 stops. With .exr or .hdr files a bit depth of 32 means that 2^32 values, 32 stops, or 4.3 billion unique values are available. However as these file types use single precision floating point numbers this is expanded to $\frac{3.4\times10^{38}}{1.18\times10^{38}}$ or 254 stops.

2.2.2 Image Acquisition

There are three primary methods for generating HDR images:

1. *Beam Splitter* cameras or *true HDR* cameras, which make use of multiple image sensors
2. *Sequential Varied Exposures*, which are merged together to form an hdr image
3. *Artificially Rendered Scenes*, which are covered later in this chapter.

Beam Splitters Beam splitter cameras, such as those developed by Doi et al. [DHKS86] and Saito [Sai95, NM00]) use a prism to split incoming light, which is projected onto multiple sensors, with different sensitivities. These images are then merged into one exposure. Recently Spheron, Panoscan and Weiss have produced commercial models of these cameras. They have the advantage of being able to produce HDR images in real time, making them ideal for video acquisition or capturing moving scenes. They are however very expensive, owing to the specialised sensors and calibration needed to produce them.

Sequential Varied Exposures A photograph taken with a fixed exposure time captures a subset of the potential range of light visible in a scene, a short exposure permits a smaller amount of light to fall on the camera sensor than a longer exposure, resulting in a darker image. Multiple images taken with varying exposure times can then be combined into a fuller representation of the observed scene. Figure 2.1, illustrates

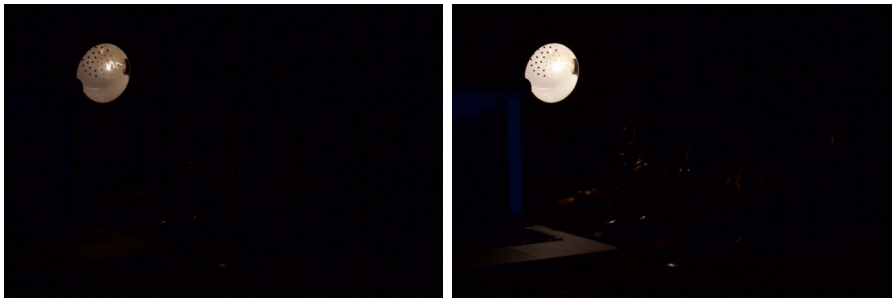
this concept visually. Mann and Picard [MP94] were the first to develop the method of combining images of various exposures in this manner. Combining images in this manner requires an accurate estimation of the camera response function which maps the brightness values captured by the camera to real world luminances. Mann and Picard's method of estimating the camera response function was however flawed, as it used a set of naive assumptions [DM08]. However, Debevec and Malik later introduced [DM08] a reliable method, that relied on image meta-data, for the approximation of the camera response function to combine multiple exposures.

2.2.3 Display of HDR images

The display of HDR images on conventional monitors requires the compression of the radiance values stored in the image to the domain which the LDR monitor is capable of representing. The group of methods for applying this transformation are Tone Mapping Operators (TMOs). A thorough review of which can be found in [BADC17]. Krawczyk [KMZS07] offers a perceptual evaluation of some simpler methods of scaling images for targeted screen brightnesses using a Parameter Estimation through Sequential Testing (PEST) method, asking participants to adjust the brightness of images such that they appear visually pleasing, the nature of the algorithm ensures that images converge to a local minimum. For an evaluation of tone mapping operators in a visual attention and eye tracking context the reader is directed to the work of Narwaria [NDSLCP14], though this work is discussed further in later sections.

The accurate reproduction of the full range of luminance stored in an HDR image requires specialised equipment. Seetzen et al. [SHS+04] developed a projection-based LCD screen, capable of displaying luminances of 0.54 to 2,700 $\mathrm{cd/m^2}$ or 12.3 stops. Dong et al. [DNPN14] utilised this display later to capture eye tracking for HDR videos.

Seetzen et al. [SHS+04] also developed an HDR monitor with an LDR front panel, which was used to control colour and limit the luminance throughput from the LED back panel, composed of 1380 individual LEDs. Their prototype screen was capable of achieving a luminance range of 0.001,5 to 3,000 $\mathrm{cd/m^2}$ [Sel13b, Muk17]. SIM2 [SIM15] later developed a number of $1080p$ monitors with an active display area of $1021 \times 572mm$,



(a) $\frac{1}{4000}$s

(b) $\frac{1}{1000}$s

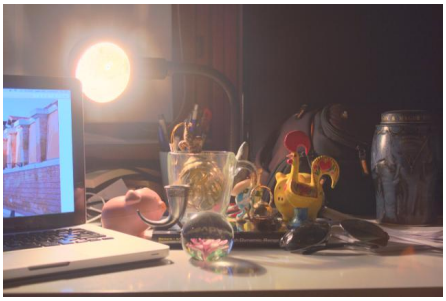
(c) $\frac{1}{125}$s

(d) $\frac{1}{30}$s

(e) $\frac{1}{8}$s

(f) $\frac{1}{2}$s

(g) 2s



(h) Tonemapped image

Figure 2.1: (a:g) Images taken with various exposure times, and (h) a tonemapped image of the combined HDR image

used in later chapters, capable of outputting luminance ranges of 0.4 to 4,000 cd/m^2, with later models reaching peak luminances of 6,000 and 10,000 cd/m^2.

2.2.4 Selective Rendering

Selective rendering can be divided into two broad categories, adaptive methods, which selectively allocate available computational resources based on some criteria, and progressive methods, which render a scene incrementally. This section focuses on the former as it bears the most relevance to this work.

Rendering is a computationally expensive process, which benefits from the efficient allocation of resources [YPG01, CCW03], an effective way to target these resources is to direct them towards salient image regions, meaning those which people have a high probability attending.

Cater et al. [CCL02] examined whether inattentional blindness could be used to reduce the quality of portions of a rendered image, in order to reduce computation time. Their experiment verified their model of task-based visual attention using eye tracking technology and revealed that users were not able to consistently notice that areas in the images that were of a reduced quality. Cater et al. [CCW03] expanded their work and combined task maps and a contrast sensitivity function to selectively reduce the number of samples in unimportant regions of an image and regions where errors would be more noticeable. This reduced the resolution of the image in less salient areas but occasionally stuttered when interpolating between neighbouring frames.

Sundstedt et al. [SCCD04] developed task importance maps, which allowed users to manually tag objects to be rendered at higher quality, which would be relevant to a given task. They determined that if sufficiently focused on a task observers would often fail to see reductions in image quality even if those low quality regions were within the foveal region. They later expanded this method [SDL+05] combining task maps with saliency maps to create overall importance maps. Longhurst et al. [LDC06] introduced a GPU-based system with live anti-aliasing within a selective rendering framework. Chalmers et al. [CDMPdS07] created low quality snapshots of scenes using rasterization and then subdivided the image into salient regions, sampling important sub-images

more frequently to achieve a higher level of perceived realism.

More recently Koulieris et al. [KDCM14] developed a system based on top down visual saliency and context based object tagging to reduce the level of detail of subsurface light transport, refraction and bump mapping in unimportant regions.

2.3 Rendering

In computer graphics, rendering is the process of simulating the traversal of light through a virtual scene, to create an image in a virtual camera. The rendering equation is a mathematical formulation of this problem introducted by Kajiya [Kaj86], relating the outgoing radiance, L_o by modelling the interaction of the incoming radiance from a solid angle at that point with the reflectance properties of the surface.
The rendering equation is the sum of the reflected, L_r, and emitted, L_e radiance in direction ω_o at point x and can thus be stated as:

$$L_o(x \to \omega) = L_e(x \to \omega_o) + L_r(x \to \omega_o) \tag{2.6}$$

The generalised reflectance function for reflected radiance L_r, given an incoming radiance, L_i, can be stated as:

$$L_r(x \to \omega_o) = \int_\Omega f_r(x, \omega_i \to \omega_o) L_i(x \leftarrow \omega_i) \cos(\omega_i) d\omega \tag{2.7}$$

where $d\omega$ denotes the differential of the solid angle, integrated across the hemisphere, Ω, and ω_i represents the incoming light vector in spherical coordinates. Equation 2.6 can thus be restated:

$$L_o(x \to \omega) = L_e(x \to \omega_o) + L_r(x \to \omega_o) = \int_\Omega f_r(x, \omega_i \to \omega_o) L_i(x \leftarrow \omega_i) \cos(\omega_i) d\omega \tag{2.8}$$

2.3.1 Reflection

When light hits a surface, portions of it are transmitted through the material, dissipated as heat and absorbed, and re-emitted. The law of the conservation of energy means

that the outgoing radiance from a point is always less than the irradiance incident at that point. The generalised Bidirectional Scattering Distribution Function (BSDF), incorporating both transmittance, through the Bidirectional Transmission Distribution Function (BTDF), and reflectance, through the Bidirectional Reflectance Distribution Function (BRDF).

A BRDF represents the proportion of irradiance incident on a surface which is reflected and describes the angular distribution of this radiance. Though the use of it is dependent on the assumption that extant radiance originates at the same location on a materials' surface as the incident irradiance it is resultant from. When this assumption can not be met, owing to the examination on a sufficiently small scale, or a translucent material, the sub-surface light transport of a material must be modelled. The Bidirectional Surface Scattering Reflectance Distribution Function (BSSRDF) models this process [JMLH01]. On a macro level it can be assumed that light is reflected from the same layers of material it 'hits', as the distance between the point of entrance and exit becomes sufficiently small. Thus the deciding factor on whether to use a BSSRDF or a BSDF is a function of scale.

The "bidirectional" part of the name is the result of the Helmholtz Reciprocity condition, that states that the function is invariant to a reversal in ingoing and outgoing directions. Thus:

$$f_r(x, \omega_i \rightarrow \omega_o) = f_r(x, \omega_i \leftarrow \omega_o) \tag{2.9}$$

BRDF models vary in complexity, the simplest of which is the diffuse or lambertian model, see Figure 2.4a, where the reflected light is constant in all directions. This can be denoted as:

$$f_r(x, \omega_i \rightarrow \omega_o) = \frac{c}{\pi} \tag{2.10}$$

where c is a constant, representing the fraction of incident light that is reflected.

Specular reflections or highlights are the directional reflections of incident light, resulting from a smooth or glossy surface. Phong [Pho75] introduced the widely used specular shader which produces a singular reflectance lobe, see Figure 2.4b for an

illustration. The model used in this thesis is an implementation by the author, adapted from the more physically accurate adaptation of the Phong model by Lewis [Lew94], which he adapted to account for helmholtz reciprocity and the conservation of energy. The implementation used in later chapters is modified such that individual specular and diffuse channels are evaluated. This is in place of the traditional method of simulating only white light and sampling the colour intersected object, in order to achieve a closer fit to the true appearance of the material.

Another class of BRDFs are the microfacet models, these model surfaces as a series of miniscule planes, oriented based on a distribution function D, these planes are considered mirrored surfaces and diffuse reflection is generated by the variation in the normals of these microfacets, an illustration of a microfacet surface can be seen in Figures 2.2 and 2.3, here the coloured lines represent the local unit normals of individual microfacets. The Walter BRDF [WMLT07], used in later chapters, is an adaptation of the widely used Cook-Torrance [CT82] microfacet reflection model, designed to simulate reflection and refraction from rough surfaces through the use of the GGX microfacet distribution function and a correction to Smith's shadowing and masking function [Smi67]. This is based on a physically plausible model of the underlying surface, and has been used to represent a wide variety of materials. Though the function simulates both reflection and transmittance, only the reflection portion is used in this study, it can be stated as:

$$f_r(x, \boldsymbol{n}, \omega_i \rightarrow \omega_o) = \frac{F(\omega_i, \boldsymbol{h}_r) G(\omega_i, \omega_o, \boldsymbol{h}_r) D(\boldsymbol{h}_r)}{4 \left|\omega_i \cdot \boldsymbol{n}\right| \left|\omega_o \cdot \boldsymbol{n}\right|} \tag{2.11}$$

where D represents the microfacet distribution function, which governs surface roughness, G represents the shadowing and masking function, F represents the Fresnel function which governs the general reflectivity of the BRDF, $\boldsymbol{n}$ represents the unit surface normal at position x and $\boldsymbol{h}_r$ is the half angle vector, which is the vector halfway between the perfect reflectance direction $\boldsymbol{r}$ and the view direction ω_o. An illustration of shadowing, whereby portions of the material are obscured by other sections of the surface. Masking is the inverse of this, and occurs when the light in the outgoing direction is blocked by the local surface. These structures aren't built into the geometry of models, and are therefore approximated by shadowing and masking functions such as Smith's function

Figure 2.2: An illustration of a microfacet surface, with vectors showing local unit normals

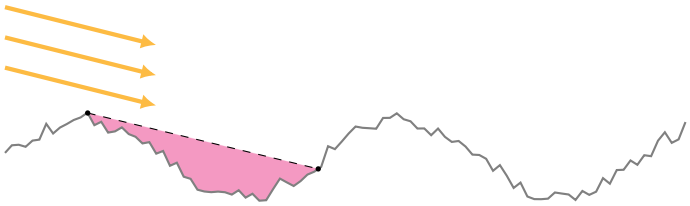


Figure 2.3: A microfacet surface, the pink shaded region indicates the area shadowed by the surface structure

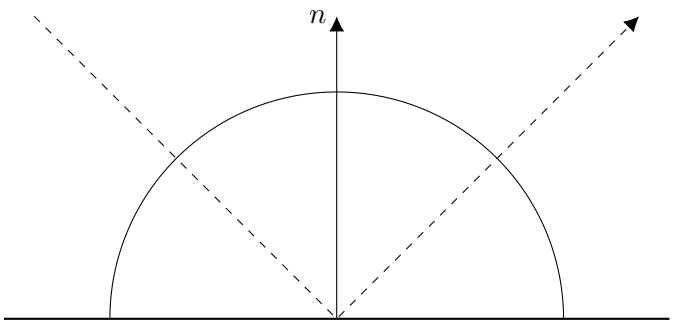
[Smi67].

2.3.2 BRDFs and Fitting

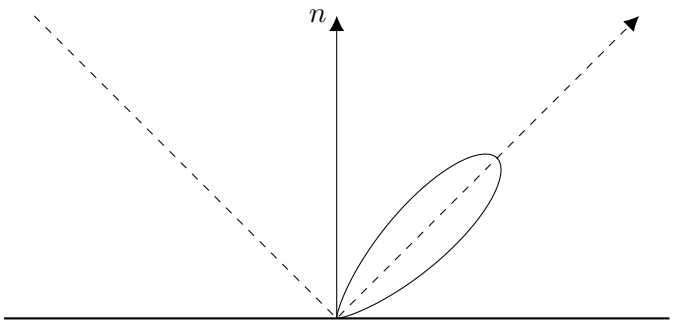
In order to reproduce the appearance of real materials digitally it is necessary to either have stored data, which represents its appearance, or determined a set of parameters for a BRDF, which will produce a function that closely matches that data. Finding parameters for representing a specific material is referred to as *fitting* that BRDF to the material. This section will focus on the literature related to acquisition of reflectance data, the approximation of reflectance using measured materials, and the fitting of BRDFs to that data.

Lafortune et al. [LFTG97] developed a system to fit multiple Phong lobes to measured materials, which proved more effective than individual lobes however they found this insufficient for some materials and found that the stability of their fitting function decreased as the number of reflectance lobes increased.

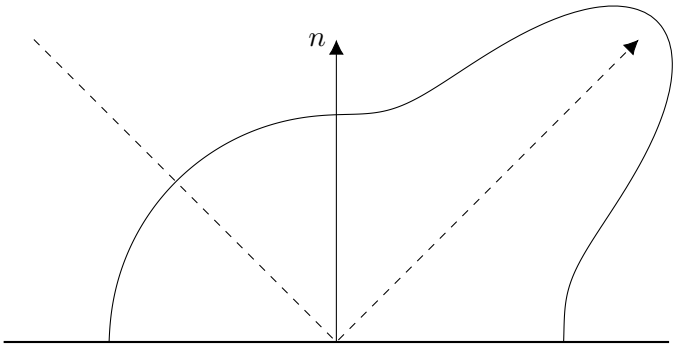
Gonioreflectometers have been employed to capture how light bounces off real world



(a) Lambertian BRDF



(b) Specular Phong BRDF, $\alpha = 18$



(c) Phong specular and diffuse BRDF, $\alpha = 18$

Figure 2.4: Simple BRDFs

materials. Dana et al. [DVGNK97] gathered sparse measurements from over 60 material samples, forming the *CUReT* database, and fitted them to the Oren-Nayar [ON95] and Koenderink [KVDS96] BRDFs.

The MERL material database [MPBM03] was the first large database of densely sampled data-driven isotropic materials, containing a variety of plastics, metals, fabrics and natural materials. It has been used as a benchmark for comparing the flexibility and accuracy of new BRDFs as a replacement for acquiring first-hand data [BSH12][LKYU12][BLPW14][HP15].

Ngan et al. [NDM05] used Sequential Quadratic Programming to minimise a squared difference error metric. They evaluated the ability of seven analytical BRDFs to represent the MERL database. The Cook-Torrance [CT82], Ashikhmin-Shirley [AS00] and He et al. [HTSG91] BRDFs fit the data well with minimal errors, while simpler Blinn-Phong [Bli77] and Lafortune [LFTG97] models performed poorly. The grazing angles were not included in the fitting, as the data was extrapolated when the MERL materials were captured. They also found that it was difficult to fit many of the materials with a single reflectance lobe. The disadvantage of using an extra reflectance lobe is the increased computation time and fitting becomes less stable.

Recently a number of BRDFs have been designed, specifically to fit to the MERL database, Bagher et al. [BSH12] introduced the Shifted Gamma microfacet Distribution (SGD) function for the Cook Torrance BRDF [CT82], replacing the traditional Beckmann distribution. Their reflectance model has a large number of parameters but can fit each material in the database with a single reflectance lobe. They precompute the values of their shadowing and masking function for offline rendering and use an approximation for GPU-based rendering as the precomputed values are slow to access in real-time applications.

Löw et al. [LKYU12] introduced the ABC BRDF which accurately models glossy surfaces. Brady et al. [BLPW14] proposed genBRDF which used genetic algorithms to generate new BRDFs from the MERL database. Holzschuch and Pacanowski's BRDF [HP15] designed a physically based BRDF incorporating both reflectance and diffraction, which provided accurate fits to the MERL database and outperformed the

SGD distribution for the Cook-Torrance BRDF, which had previously provided the best fits to the MERL database.

Chapter 3

Visual Attention and Eye Tracking

3.1 Introduction

This chapter details the background theory and literature relevant to visual attention, eye-tracking studies and work on visual fatigue.

3.2 The Human Visual System

The Human Visual System (HVS) is composed of the eye, the surrounding nerve system and the section of the brain that processes visual sensory information. A graphical represention of the eye can be seen in Figure 3.1, which was originally from the work of Harvey [Har11]. The eye has three layers, the sclera, the chloroid and the retina. Light entering the eye passes through the cornea and pupil through the pupil, its amplitude modulated by the iris and focused by the lens onto the retina. The retina covers the inner surface of the eye, and contains photoreceptor cells which are sensitive to light. These photoreceptors can be divided into two categories; rods and cones. Cones are sensitive to light in the photopic range ($10^{-2} - 10^{6}\mathrm{cd\,m^{-2}}$), and are sensitive to colour, high spatial frequency changes in illumination and fast changes in light intensity [BADC17]. In contrast, rods are sensitive in the scotopic range ($10^{-6} - 10^{2}\mathrm{cd\,m^{-2}}$), are

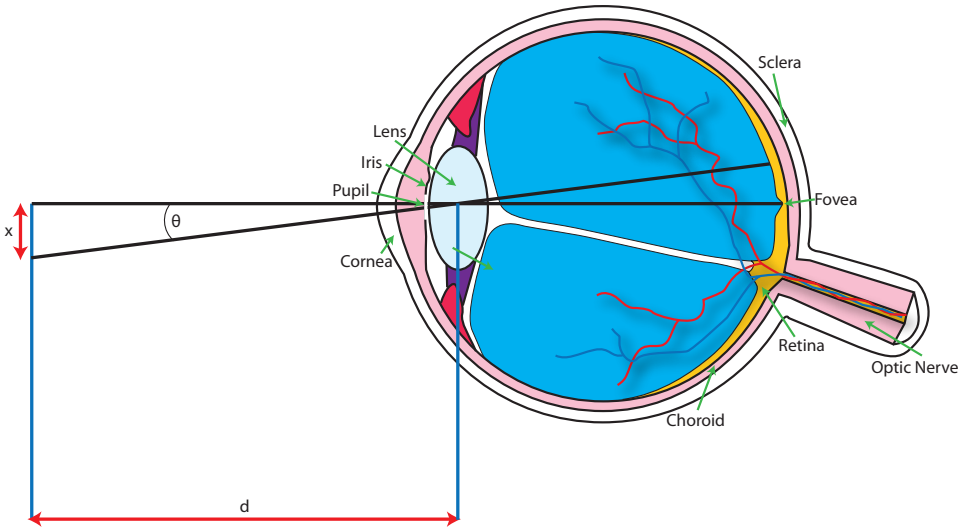


Figure 3.1: A diagram of the human visual system, used with permission from [Har11]

insensitive to colour and high spatial frequency variations. Rods are also not present in the fovea, which possesses the greatest concentration of cones [BADC17]. From here visual information travels to the vision centre of the brain through the optic nerve.

3.2.1 Visual Fatigue

Visual fatigue, eye-strain or asthenopia is a feeling of discomfort around the eyes, the leading cause of which is 'dry eye' [BVK+05]. The term computer vision syndrome was coined for this affect, as visual fatigue is a proven consequence of prolonged interaction with computer monitors [Cos94, KS01, She92, YT89].

Recent work examining the effects of luminance and ambient illuminance while viewing e-readers by Benedetto et al. [BCDZ+14], determined that these factors have an effect on the visual fatigue experienced by viewers. Sheedy et al. [SSH05], examined ambient illumination in the context of LDR displays and visual fatigue, and found significant results. Most recently Yang et al. [YJL17] examined the effect of OLED and LCD HDR displays on visual fatigue, by asking participants to determine the orientation of a grating target, see Figure 3.2, while displaying white circles of varying luminance

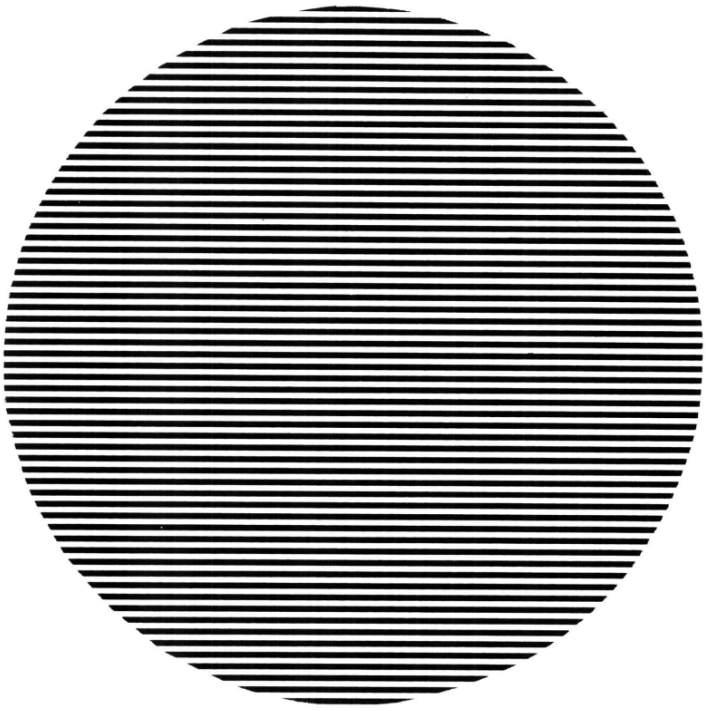


Figure 3.2: A grating target, which can induce visual fatigue

to assess brightness discrimination thresholds, and found a positive correlation, though the maximum luminance they examined was 1,568 $\mathrm{cd/m^2}$.

3.3 Visual Attention

Visual attention describes the process by which the awareness of an observer is directed, it is a subset of the study of selective attention, which incorporates the other senses. The human visual system's ability to process information is limited [BI13, KMS+06], with only a fraction of visual information which is incident on the retina being processed, and thus attention is limited in unattended locations. Visual attention can be separated into two processes bottom-up, involuntary, or exogenious, visual attention and top-down, voluntary, or endogenous, visual attention [Jam13]. Bottom-up models identify regions of interest based on assumptions of the structure and nature of the Human Visual System [BI13, Not05] and certain features or objects which are known to draw attention, including preattentive features, a good summary of which can be found in Healy and

Enns' paper [HE12].

Top-down attention is task focused, governs the direction of attention in a task-focused manner, influenced by the objective that the observer is given. This was first demonstrated in experiments by Yarbus [Yar67] in which observers were asked to view a painting by Ilya Repin, see Figure 3.3, and cued with different objectives before viewing.

Posner [Pos80] observed that attention could be shifted voluntarily to the peripheral vision, this is termed covert attention. Overt visual attention is the process by which attention is directed by bringing the fovea to focus on areas of interest, which is observable with an eye-tracker. Mechanisms of covert visual attention govern the assignment of attention to conspicuities in the periphery of vision. This process can be either voluntary or involuntary, Posner [Pos80] noted that attention could be consciously directed to specific points in his peripheral vision. Overt attention, as previously stated, is directed by fixations. Between fixations fast eye movements occur, called saccades, though attention is suppressed during this period [Mat74]. A thorough review of the theory behind saccades and fixations is given by Rayner [Ray98].

3.3.1 Saliency

The computational model of visual attention, upon which most modern visual attention models are based, was first proposed, though not implemented, by Koch and Ullman [KU87]. Their work took inspiration from the neuronal structure of primates.

Saliency maps are the output of these models, represented as greyscale images where each pixel value corresponds to how likely an observer is to pay attention to that point in the scene, the higher the pixel value the higher the saliency of that point.

Itti et al. [IKN98] implemented the ideas of Koch and Ullman [KU87] and developed a computational model of bottom-up visual attention, using a scale-space feature detection method on an image, subsampling and smoothing the image to emphasise conspicuous features. Their method performed well on a variety of images but struggled in images with uniform noise. Though this model is used widely throughout literature, even being used as a benchmark against which to test high-fidelity models of attention [BPT10] it is highly inaccurate, scoring next to last on the MIT saliency benchmark

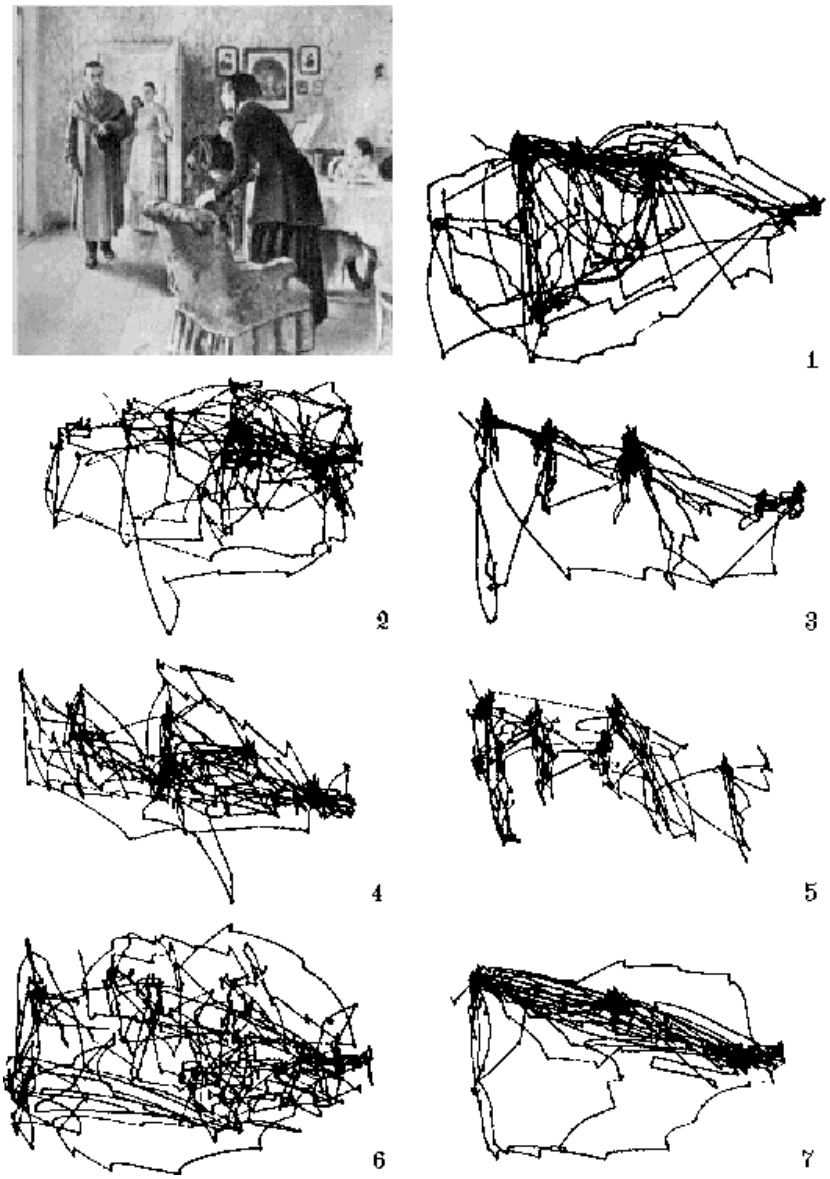


Figure 3.3: Eye tracking data recorded by Yarbus [Yar67] for a single participant provided with seven distinct tasks, the viewing time for each task was three minutes. The tasks are as follows and repeated verbatim from the original translation, for accuracy: 1) Free-examination of the picture; 2) estimate the material circumstances of the family in the picture; 3) give the ages of the people; 4) surmise what the family had been doing before the arrival of the 'unexpected visitor; 5) remember the clothes worn by the people; 6) remember the position of the people and the objects in the room; 7) estimate how long the 'unexpected visitor' had been away from the family.

[BJB+], above only random chance, and the Achanta et al. [AHES09] model, which is primarily used for object detection. Harel et al. introduced the Graph-Based Visual Saliency (GBVS) model [HKP06], which used normalized activation maps and graph cuts to identify salient regions within an image, their method combines well with other methods. Judd et al. [JDT12] noted that GBVS' strong performance is likely a result of its strong centre bias, as observers fixations tend to exhibit a centre bias as well. GBVS is an effective bottom-up map, scoring relatively well but not exceeding the predictive power of the best deep learning models [KTB14, JHDZ15], which have become the norm in recent years.

3.3.2 Visual Attention in High-Fidelity Applications

Through necessity visual attention models have been recently adapted for high-fidelity imaging techniques. Models for stereoscopic displays have recently emerged, the complexity of these models is higher than conventional methods as depth must be added as a new feature. Wang et al. [WFN+14] introduced a new 3D saliency model for image retargeting, their method estimated saliency using 2D feature maps and accounted for depth by incorporating a model of stereoscopic bias. Banitalebi et al. [BDPN16] introduced a learning based model for 3D saliency, utilising an eye tracking experiment involving 24 participants for evaluation. Fang et al. [FLL+17] proposed a machine learning model for estimating stereoscopic saliency, their method used traditional 2D cues along with a depth map and semantic top-down information to learn a saliency model that performed well compared to previous stereoscopic models.

Rai et al. [RGLC17] conducted an eye tracking experiment, obtaining fixation data for 60 360° videos from 40 participants, for models of omnidirectional saliency. They offer a review of these methods in their paper.

Narwaria et al. [NDSLCP14] offered an assessment of the effect of applying various TMOs to HDR content on the viewing strategies of observers. They utilised a screen with a peak luminance of 4,000 $\mathrm{cd/m^2}$, and while noting that the ITU-R BT.500-13 recommendation [IR12] used to establish the HDR MPEG classification [LFH15] recommends an ambient luminance of 600 lx in this scenario (0.15 of maximum luminance),

their experimental environment was not equipped to meet this specification.

3.4 Metrics for Saliency Map Comparison

Bylinkskii et al. [BJO+16] offered an evaluation of eight different common metrics for measuring the performance of saliency maps. They attempted to offer an intuitive explanation of the inner workings of each. It is their implementation of these metrics, for use in the MIT saliency benchmark [BJB+], that was used in this thesis, and their code for metric visualisation was used to generate the visualisations in this section.

3.4.1 Area under the receiver operating characteristic curve

The Area Under the receiver operating characteristic Curve, commonly referred to as the AUC metric, is the most widely used method for comparing a saliency map against a set of fixations [BJO+16]. The Receiver Operating Characteristic (ROC) curve emerged from signal detection theory, representing the relation between the true and false positive rates of a binary classifier function [Nev69]. The term *true positive* refers to a prediction or observation on a condition which is found to be positive, where reality matches that result, i.e. the condition is positive and is found to be as such. The term *false positive* refers to a result which is found to be positive, despite the condition being negative in reality. Table 3.1 illustrates this visually, including the conditions for false and true negative predictions. In predicting fixations, a false positive is when a saliency map predicts that a region will be fixated, when it is not actually fixated upon. A true positive represents the correct prediction of fixation locations. The ROC curve is built by adjusting the sensitivity of the binary classifier and plotting the true positive rate against the true negative rate. For saliency maps, this sensitivity adjustment is achieved by choosing a series of pixel value thresholds, chosen by the value needed to incorporate a fixed percentage of fixations. The resulting threshold maps operate as binary classifiers, with pixel values above the threshold indicating a predicted fixation, or a predicted positive, and pixel values below the threshold indicating a region that is predicted to not be fixated. This is best represented visually, Figure 3.4 shows an image from the Fairchild database [Fai07], threshold maps for the automatically selected

	Actual Positive	Actual Negative
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

Table 3.1: A representation of the meaning of true and false positives

values and an illustration of true and false positives. Here green values indicate a true positive and red values indicate a false positive. The resulting ROC curve is displayed in Figure 3.5. The code for this visualisation comes from the MIT Saliency Benchmark [BJB+], and the work of Bylinskii et al. [BJO+16].

Though this metric has uses outside visual attention, Hanley and McNeil offer an intuitive review of the use of the AUC metric for medical applications [HM82], whereas Bradley [Bra97] and Fawcett [Faw06] offer reviews of its applications in machine learning.

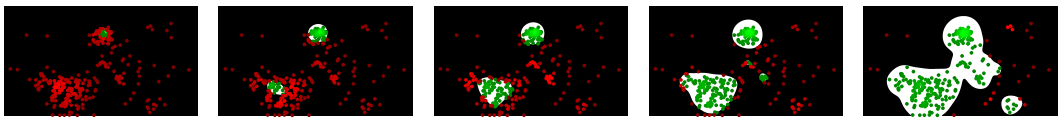
3.4.2 Earth Mover's Distance

The Earth Mover's Distance (EMD) was originally introduced by Rubner et al. [RTG00] as a method for retrieving images from large databases, commonly referred to as image retrieval. It is a probabilistic model that measures the cost of transforming one distribution into another, this cost is the amount of data that is to be moved and the associated distance that it must be moved. This is akin to the cost of moving one pile of bricks, so that it is in the same configuration as another distinct pile. The more similar the original piles, the lower the cost incurred in moving them. The algorithm searches over the space of possible transformations, and converges to a global optimum. This is incredibly computationally expensive, especially for large images. Pele and Werman [PW08] developed a robust version of the algorithm, which operates in linear time and introduced a reliable ground distance measure for Scale-Invariant Feature Transform (SIFT) matching. They later built upon this work [PW09], imposing limits on the distance that information can be moved to reduce the space of possible solutions. It is the re-implementation of their algorithm by Bylinskii et al. [BJO+16] that is used in this thesis. Figure 3.6, shows a visualisation of this metric, also utilising code from

(a) Original image [Fai07]

(b) Fixation density map

(c) Threshold images



(d) Visualisation of true and false positives

Figure 3.4: A visualisation of the AUC metric, code for visualisation provided by Bylinskii et al. [BJB+, BJO+16]

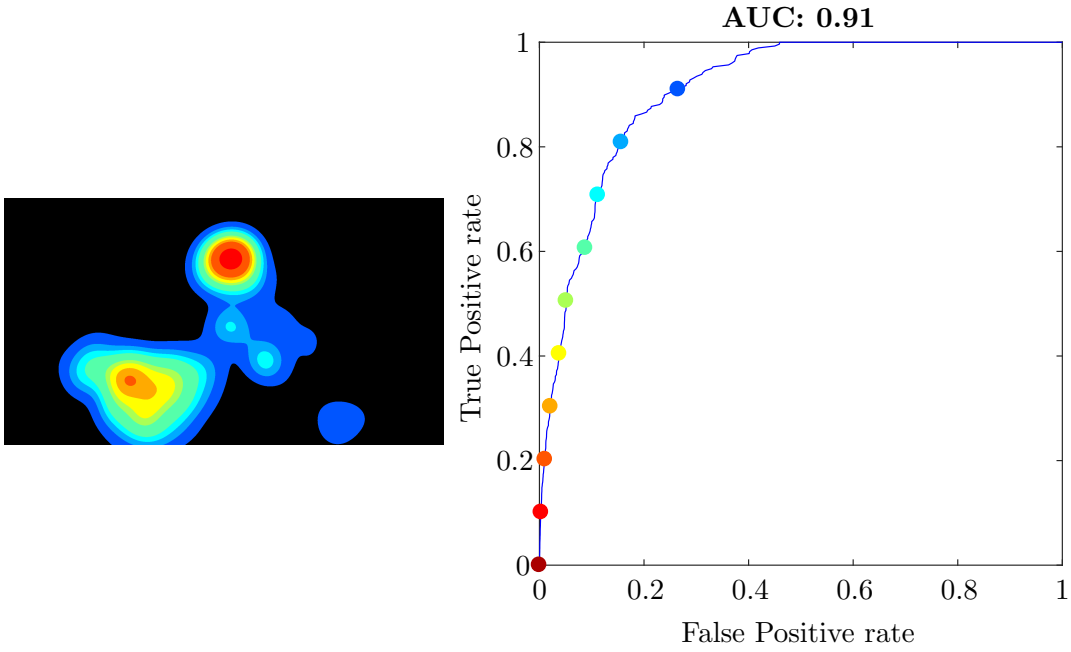


Figure 3.5: The overlaid threshold maps from Figure 3.4 and a graph showing the resulting ROC curve for comparing a fixation density map against fixations from human participants

Bylinskii et al., for fixation density maps from two sets of participants collected later in this thesis. Here the red areas represent areas from which information was moved, whereas the green regions represent areas to which information was moved.

3.4.3 Similarity Metric

The SIMilarity (SIM) metric was introduced by Swain and Ballard [SB91], as histogram intersection, as a method for improving object detection and localisation for autonomous robots by fast colour histogram matching. The takes in two probability mass functions as inputs, that is the sum of both maps must be equal to one. The Similarity score, S, of the two distributions P and Q is then measured as:

$$S = \sum_{i}^{n_{px}} \min(P_i, Q_i) \tag{3.1}$$

where

$$\sum_{i}^{n_{px}} P_i = \sum_{i}^{n_{px}} Q_i = 1 \tag{3.2}$$

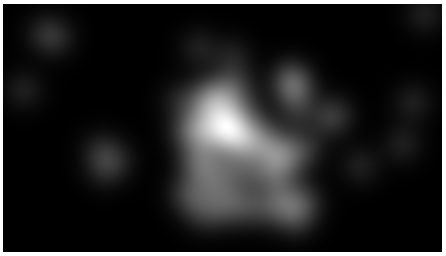
where i represents the i^{th} pixel and n_{px} is the total number of pixels. A visualisation of the dissimilarity between the fixations of two groups of participants is shown in Figure 3.7. The nature of this measure results in a heavy penalisation of offsets, as only exact locations are compared, rather than local regions.

3.4.4 Pearson's Correlation Coefficient

Pearson's Correlation Coefficient (CC) measures the linear correlation of two probability distributions by comparing their covariance, this is a symmetric measure of the manner in which these probability functions vary with each other. A score of one would indicate a perfect positive correlation, and a score of zero a perfect negative correlation. To calculate the score C, first the saliency or fixation maps P and Q are normalised such that:

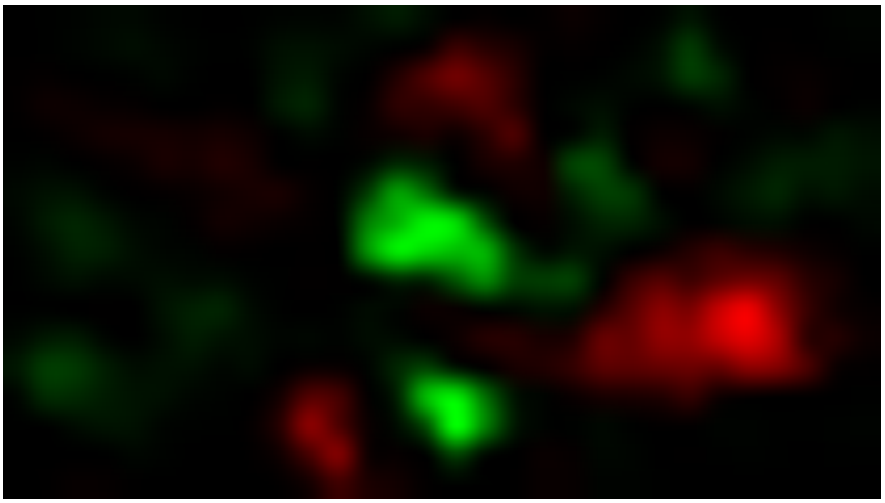
$$\hat{P} = \frac{P - \mu_P}{\sigma_P} \tag{3.3}$$

(a) Original Image [Fai07]



(b) Fixation density map 1

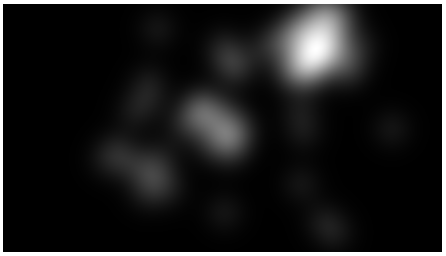
(c) Fixation density map 2



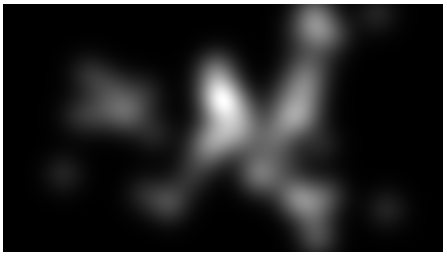
(d) Visualisation of the Earth Mover's Distance

Figure 3.6: A visualisation of the Earth Mover's Distance between fixation density maps captured from two groups of participants

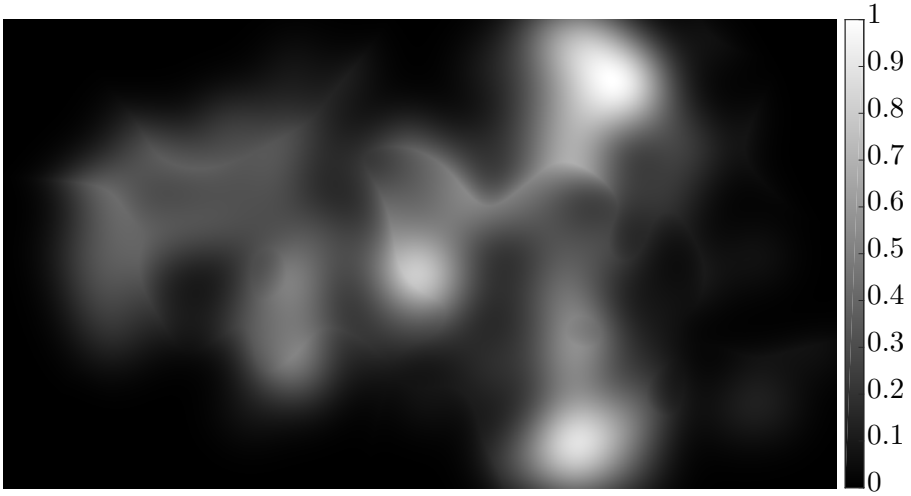
(a) Original Image [Fai07]



(b) Fixation density map 1



(c) Fixation density map 2



(d) Similarity Map

Figure 3.7: A visualisation of the Similarity score between fixation density maps captured from two groups of participants

and

$$\hat{Q} = \frac{Q - \mu_Q}{\sigma_Q} \tag{3.4}$$

where $\hat{P}$ and $\hat{Q}$ represent the normalised maps, μ_P represents the mean of all pixels in map P, and σ_P their standard deviation. The correlation coefficient can then be calculated as:

$$CC(P,Q) = \frac{\sum_i^{n_{px}} (\hat{P}_i - \mu_{\hat{P}})(\hat{Q}_i - \mu_{\hat{Q}})}{\sigma_{\hat{P}}\sigma_{\hat{Q}}} \tag{3.5}$$

A visualisation of this similarity for an image from the Fairchild database [Fai07] is shown in Figure 3.8. Here lighter areas indicate areas of higher correlation.

3.4.5 Normalised Scanpath Saliency

Peters, along with Itti and Koch [PIIK05] developed the Normalised Scanpath Saliency (NSS) to measure the accuracy of saliency models to predict human fixations on an image. The metric operates by evaluating the values of a normalised saliency map at fixated locations in image space. The metric score can thus be calculated as:

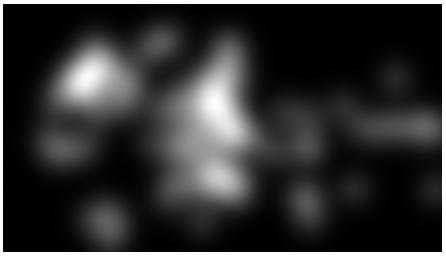
$$NSS(P,Q^b) = \frac{1}{\sum_i^{n_{px}} Q^b} \sum_i^{n_{px}} \hat{P} Q_i^b \tag{3.6}$$

where Q^b is the binary fixation map with fixated locations having a value of one and non-fixated locations a value of zero, that the saliency or fixation density map P is evaluated. Figure 3.9 shows a visualisation of the NSS metric applied across the fixations of two groups of participants for an image from the Fairchild database [Fai07].

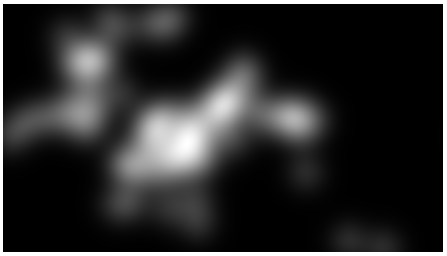
3.4.6 Kullback-Leibler Divergence

Kullback-Leibler Divergence (KLD) has its base in probability and information theory [Kul97], as a measure of the cross-entropy of two probability distributions. In other words it represents the cost of moving from one probability distribution to another. In codeword length optimisation [BRY98] the intuitive understanding often cited is that $KLD(P,Q)$ represents the cost incurred by encoding distribution Q with codewords

(a) Original Image [Fai07]



(b) Fixation density map 1

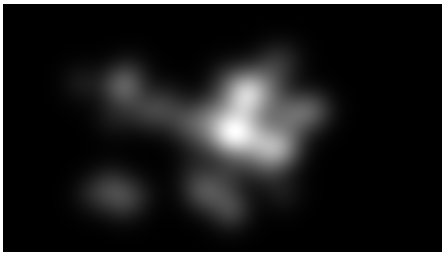


(c) Fixation density map 2

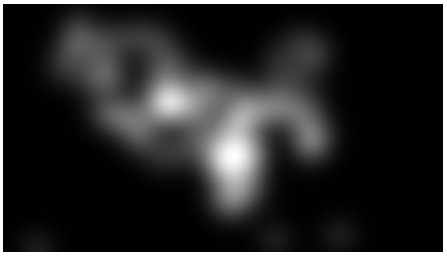
(d) Correlation map

Figure 3.8: A visualisation of Pearson's Correlation Coefficient between fixation density maps captured from two groups of participants

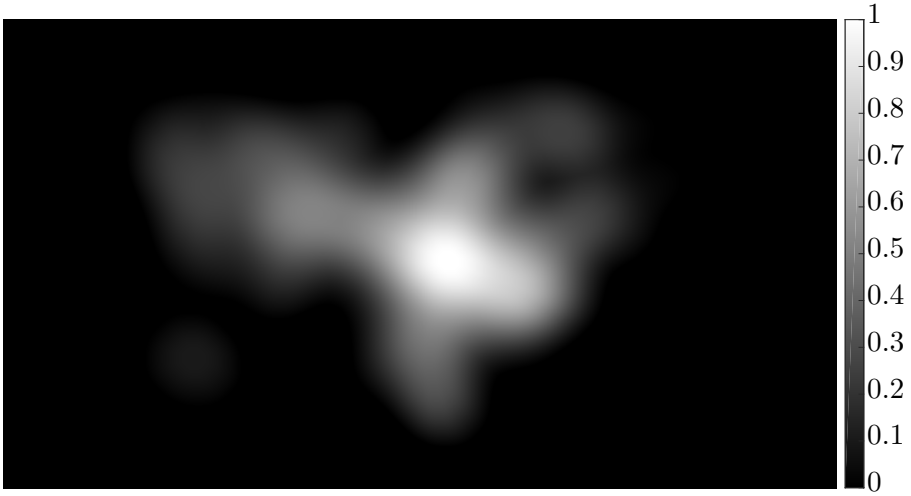
(a) Original Image [Fai07]



(b) Fixation density map 1



(c) Fixation density map 2



(d) Normalised saliency

Figure 3.9: A visualisation of Normalised Scanpath Saliency between fixation density maps captured from two groups of participants

optimised for distribution P [GBC16]. Though the KLD is not a true metric [GBC16], as it is not symmetric, as $KLD(P,Q)! = KLD(Q,P)$. The implementation used in this paper is a symmetric application of the code available in the MIT Saliency Benchmark [BJB+, BJO+16] and is computed as follows:

$$KLD(P,Q) = \frac{1}{2}\left(\sum_{i}^{n_{px}} Q_i \ln\left(\varepsilon + \frac{Q_i}{\varepsilon + P_i}\right) + \sum_{i}^{n_{px}} P_i \ln\left(\varepsilon + \frac{P_i}{\varepsilon + Q_i}\right)\right) \tag{3.7}$$

where ε is a regularisation constant [BJO+16]. A visualisation can be seen in Figure 3.10.

3.4.7 Summary

The above metrics are used in this thesis for a variety of reasons. While all are used in the MIT Saliency Benchmark [BJB+], the Area Under the Curve and Normalised Scanpath Saliency metrics are the current metrics by which the models are benchmarked. Pearson's Correlation Coefficient has a historical significance as it is used in many statistical tests. The similarity score, while a relatively naive algorithm has been used in recent HDR studies [NKHE15]. The Earth Mover's Distance and Kullback Leibler Divergence metrics, while not specifically designed for fixation or saliency map comparison have strong roots in probability and information theory and offer a more probabilistic perspective on the comparisons that appear in later chapters.

3.5 Eye tracking

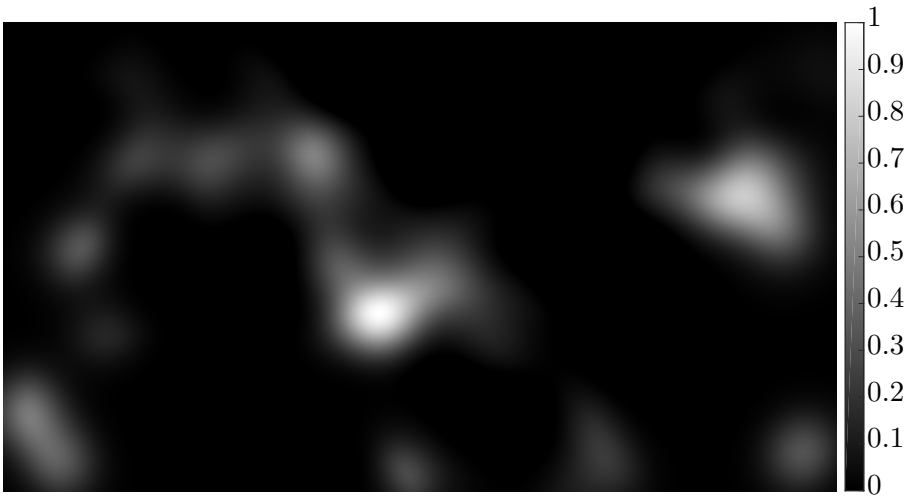
Eye tracking describes the process of estimating the position and orientation of a person's eyes to predict the direction of their gaze, or their point of focus, through the use of specialised hardware and software. While these technologies serve as invaluable resources for the verification of models of overt attention, their applications stretch well beyond the validation of visual attention models.

The first person to examine the eye-movements of observers was Javal [Jav78] in 1879, who observed the eyes of subjects while reading and noticed that their movements were not fluid, instead being constituted by fast motions, which he termed 'saccades',

(a) Original Image [Fai07]

(b) Fixation density map 1

(c) Fixation density map 2



(d) Divergence map

Figure 3.10: A visualisation of Kullback Leibler Divergence between fixation density maps captured from two groups of participants

and fixations. These observations were recorded by hand. Early efforts at creating eye tracking devices were varied, inaccurate, and had the potential to cause significant discomfort to the viewer.

Early methods could be roughly divided into four approaches; manual observation, specialised contact lenses, mechanical, and elecro-oculography. Javal's [Jav78] work, was the first manual observation method, he stood behind participants and observed their eyes using a mirror, noting the eyes' movements. This method has obvious drawbacks and is prone to large measurement errors. This work was later built upon by the refinement and use of optical instruments that provided greater magnifications [New28, GTL23] and even using stereoscopes to achieve a greater fidelity [Pec34]. Barlow [Bar52] utilised the direct application of mercury to the cornea, examining the movement of a small droplet using a camera to measure eye movements. However these methods are not in common use today. Electro-oculographic methods of measuring eye movements, detect rotations of the eye by measuring changes in the skin around the eye using electrodes. Schott [Sch22] was among the first to use these methods but they are still in use today in the field of human computer interaction [BRT08].

In 1908 Huey [Hue08] developed the first contact lens method, attaching opaque contact lenses with a small hole positioned over the pupil to participants' eyes, a pointer was attached to the lens which would allow for easier recording of the gaze direction. He used this device to track the participants' eye movements while reading and could identify the words on which the participants' gaze lingered. Marx and Trendelenburg [MT11, Yar67] attached mirrors to aluminium cups which were then affixed to the eye, requiring anaesthetic to reduce discomfort. This was one of the early reflection based methods, whereby researchers [Doh25, RR50, DG52, Nac59] would shine beams of light on mirrored cups or contact lenses attached to the eye. Adler [AF34] took a more invasive route and attached the mirror directly to the sclera of participants. These methods were relatively accurate for the time [Yar67] but offer suffered from several disadvantages; first, they require the direct attachment of apparatus to the eye; second, stemming from this their mass potentially affects the movements on the eye; third, when not attached securely to the eye saccades could cause shifts in their position resulting

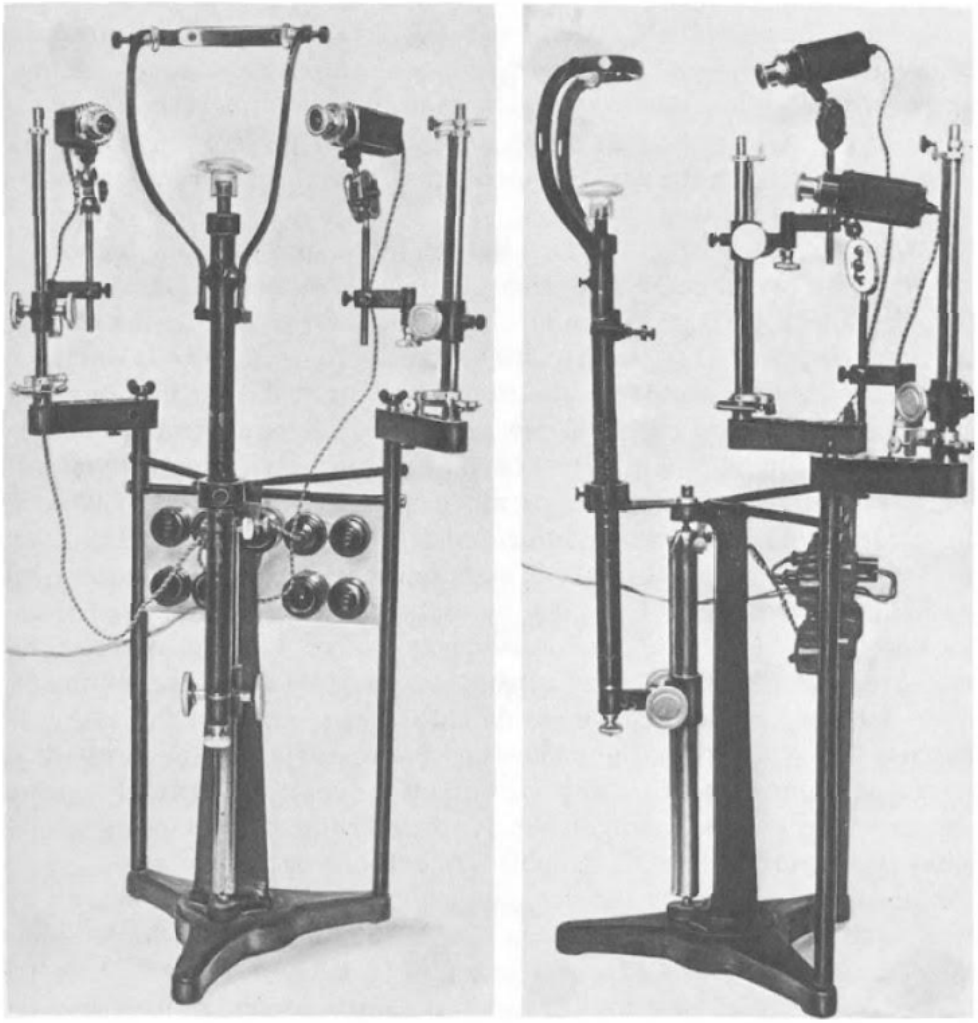


Figure 3.11: An image of the eye tracking device employed by Yarbus [Yar67]

in measurement error [Yar67]; and finally, the majority of cases required that the eyes be anaesthetised.

The most notable mechanical method was that of Yarbus [Yar67], whose work on task dependent visual attention formed the foundation of modern research on top down attention. Yarbus' apparatus, shown in Figure 3.11, utilised suction cups attached to the eye-balls to measure eye movements. While invasive this method was effective, the results of one such task can be seen in Figure 3.3.

Cornsweet and Crane [CC73] developed what would become the bases for table-mounted eye tracking devices commonly used today. Using infra red cameras to capture

reflections from different layers of the eye, called Purkinje images, to estimate the orientation of the eye. These methods have become common place, Rayner [Ray98], Duchowski [Duc02] and Morimoto et al. [MM05] offer reviews of these methods.

3.5.1 HDR Visual Attention and Eye tracking

Dong et al. [DNPN14] recorded eye-tracking data for ten HDR video sequences, using 18 participants. They utilised the projection display developed by Seetzen et al. [SHS+04], which has a peak luminance of 2,700 cd/m^2.

Nasiopoulos et al. [NDK14] later used the same equipment to contrast viewing strategies of HDR and LDR videos, they found that when viewed separately there was no significant preference for HDR or LDR displays, though when viewed simultaneously there was a significant preference. Their experiment found no significant difference in eye-movements for either condition.

Petit, Brémond and Tarel [BPT10] [PBT09] derived a *Contrast Features* (CF) saliency map for HDR images based on that of Itti and Koch [IK00] by normalising the intensity feature and adjusting the orientation feature so that it is invariant to contrast. They conducted an eye-tracking experiment in which participants were asked to look at a real HDR scene, created using a dark room, a box and several high-intensity lamps. Testing their saliency map against this tracking data demonstrated that for this scene their saliency model outperformed the Itti and Koch model.

Work by Nemoto et al. [NKHE15] used an HDR monitor with a peak brightness of 4,000 cd/m^2 to examine how visual fixation patterns and intensities varied between HDR and tonemapped LDR images. They focused on differentiating the effects of HDR and LDR stimuli using the similarity metric. Rather than enforcing a uniform dynamic range across images the method of Krawczyk et al. [KMZS07] was used to scale the images to fit the range of the monitor, though this method was based on tests performed on an LDR screen where the brightness did not exceed $250nt$. Their HDR stimuli had dynamic ranges from < 48dB to > 84dB so it is difficult to conclude the effect of screen brightness on dynamic range from their work. Though they found that in half the examined images that the fixation patterns and/or dwell time varied significantly

between HDR and LDR stimuli.

Banitalebi-Dehkordi et al. [BDDPN15] introduced a "Learning-Based Visual Saliency fusion method for HDR content" (LBVS-HDR), which utilised Random Forests in its model. They trained their model using eye-tracking data from ten HDR videos, viewed by eighteen observers on a 2,700 cd/m^2 projection-based screen. Their model outperformed other learning-based saliency methods across a variety of metrics.

Dong et al. [DPN16] developed a new bottom-up visual attention model for HDR images and videos, loosely based on the Itti and Koch's model. Their model requires some additional inputs, such as viewing distance and screen size, and accounts for the non-linearity of the human visual system's response to luminance, based on absolute intensity and relative to the spatial frequency and amplitude of luminance variations. Results of their eye-tracking experiments, using twenty-three HDR images and ten HDR video sequences, show a higher average result across three metrics (AUC, CC, symmetric KLD) for their method. However it isn't stated in their paper whether this improvement is statistically significant.

Chapter 4

Visual Attention for High-Fidelity Imaging

This chapter builds on Chapters 2 and 3, which covered recent developments in selective rendering and HDR eye tracking studies, and explores the areas in these fields that require further development. The following three chapters detail the collection and development of the bulk of this thesis' contributions.

The structure of this chapter is as follows. The first section presents an overview of the work in this thesis, and contextualises it in a selective rendering context, the second explores the current state of the art, and summarises current areas that require further work. The subsequent section builds on this, stating the research questions and outlining the key points this work aims to address. Finally, the last section details how these objectives will be approached in the coming chapters.

4.1 Introduction

Visual attention is a broad field, incorporating the study of both what draws a viewers conscious focus [KU87], the limitations of the human visual system [Mat74], and the manner in which attention is directed [Pos80]. These sets of techniques have broad applications for standard imaging techniques, shaping the design of road safety signs, websites, user-interfaces, camouflage, advertisements and object placement in video

games. The recent emergence of high-fidelity imaging techniques, which present new challenges, environmental factors, and costs has opened up a new avenue of research in the field of visual attention. Steps have already been made towards adjusting models of saliency for stereoscopic images [FLL+17], 360° video [RGLC17], and HDR images [BPT10]. However, the underlying experimental data, needed to form or verify visual attention and saliency models, is lacking for high-fidelity modalities. This thesis attempts to address this in the following chapters.

Targeting selective rendering, in a manner which minimises the drop in perceived quality of the rendered image requires an accurate visual attention model. If the employed model is inaccurate the reduction in quality through which selective rendering methods save resources may mistakenly reduce the image quality in a visually important region. As can be seen from the analysis of existing literature in Chapters 2 and 3 and the next section, while many frameworks exist for selective rendering with traditional imaging, the data required to create an adaptive HDR saliency model is currently insufficiently available. This thesis seeks to supplement existing selective rendering methodologies by both building upon the existing methods of capitalising on the limitations of the HVS to reduce rendering times and developing the tools necessary to adapt models of visual attention to high fidelity image synthesis pipelines, such as those of HDR content generation and physically-based rendering. This tailoring of visual attention models to high fidelity imaging will enable a higher quality rendered image, and minimise the misallocation of computational resources in a medium for which few models exist, at the time of writing.

4.2 The current state of the art

4.2.1 Selective Rendering

Rendering, ray tracing in particular, is a computationally expensive process. As detailed in section 2.2.4, many works have developed methods of exploiting the limitations of the human visual system to alleviate a fraction of this cost.

In recent years, a number of selective rendering methods have emerged. Previous

(a) Image '507' from the Fairchild Database [Fai07]

(b) The corresponding fixation map

Figure 4.1: (a) An example image and (b) its corresponding fixation map

work has focused on a number of resource management techniques, a few examples are: the direction of samples [YPG01, CDMPdS07], the reduction of resolution [CCL02], or the reducing the number of polygons in the scene [BCP03]. These methods achieve varying levels of success, with some providing significant computational speed-ups. However, not one seeks to minimise the computational load of calculating reflection. Physically-based ray tracing calculates reflection using BRDFs, which vary widely in complexity. From models such as the Phong BRDF [Pho75], which consists of only a few multiplications, to more complex microfacet models which incorporate refraction, such as the Walter BRDF [WMLT07], which require the evaluation of multiple functions.

Thus if the limited attentional bandwidth of the human visual system can be exploited to reduce the complexity of BRDFs in less salient regions, a new tool for aiding in the reduction of computational costs would be available. This however needs to be verified with psychophysical tests.

4.2.2 HDR Visual Attention

There is some debate over the assumptions about the neurological structure of the visual system that visual attention models make [BI13]. Therefore validating attention models with real world data is essential, for overt visual attention the requisite data takes the form of eye tracking datasets. The most widely-used system for evaluating the performance of bottom-up visual attention models is the MIT saliency benchmark

[BJB+], which contains both a LDR test dataset and a framework for model assessment.

The availability of testing data for HDR content is somewhat limited. Dong et al [DNPN14] created a dataset for HDR videos, using eighteen participants, later Nasiopoulos et al [NDK14] used this dataset to test user preference for LDR and HDR video. Their study also found no significant difference between viewing patterns for LDR and HDR. Though their sample size is somewhat limited and dynamic range was not uniform across stimuli. Brémond and Petit [BPT10] utilised eye tracking to validate their contrast features model, however they used only one real scene, displayed in a lit box rather than on an HDR display. Dong et al [DPN16], in a similar vein to earlier work by Nasiopoulos et al [NDK14] evaluated a learning based model of visual attention, using eye-tracking data from their dataset. They evaluated the performance of their model using three metrics (AUC, Pearson's Correlation Coefficient and Kullback Leibler Divergence) and while their results show a higher performance for their HDR saliency map they don't report whether this was statistically significant.

Furthermore, in the above papers, it isn't stated whether the LDR Itti and Koch model [IKN98], which while ground-breaking and widely used is a particularly poor model compared to the current state of the art, was adjusted for the linear encoding of brightness of HDR images. It stands to reason that a model designed for 8-bit gamma-encoded LDR images would perform poorly on linear HDR images with higher bit-depths. This fact bears little relevance to whether HDR models of visual attention are necessary, past accounting for a difference in luminance encoding and difference in just noticeable difference (JND) levels.

The work of Nemoto et al [NKHE15] aimed to measure the effect of HDR content on visual attention, while their method compares LDR and HDR fixations they rely entirely on two metrics, Similarity and Kullback Leibler Divergence, which compare blurred fixation density maps, an example of which can be seen in Figure 4.1, and omit the gold standard area under the curve (AUC) metric. More importantly their work didn't control for the peak brightness of the image in their analysis or dataset. In addition the LDR versions of the image were scaled using a method [KMZS07] designed to maximise perceived quality for tonemapped images, rather than enforce a uniform

brightness.

4.2.3 Eye strain

While research has been conducted on the effects of digital readers [BCDZ+14], 3D televisions [LHP10, LFI+10] and LDR visual display terminals [KS01, LPWM09] on visual fatigue, and results show that an increase in brightness is correlated with an increase of eye-strain [BCDZ+14, YJL17]. No research has been done to examine this effect with HDR displays or its interaction with visual attention. This is important as uncomfortably bright image regions may have an effect on visual attention.

4.3 Research Question and Objectives

Reducing inefficiencies in selective rendering can be done in a number of ways, this thesis focuses primarily on two: developing a new way to selectively allocate resources which is capable of interacting with existing technology and examining if existing visual attention algorithms need to be adapted for unexamined variables.

Establishing a firm basis for the evaluation, and need for, HDR models of visual attention requires two things: a reliable HDR eye tracking dataset for validation, and experimental evidence that a need does or does not exist for expanding the manner in which luminance is processed in these models, i.e. is luminance a feature? While Nemoto et al [NKHE15] and Dong et al [DNPN14] have collected HDR eye tracking data, these datasets are relatively small and dynamic ranges are too varied for a statistical analysis of the effect of high luminances on visual attention. Addtionally while some research contrasts LDR and HDR visual attention [BPT10, NDK14, NKHE15] none adequately tests for the effect of brightness on visual fatigue, either due to insufficient data, or a lack of control for brightness. Thus an area of research appears which needs to be tackled before meaningful models can be developed. Namely, the effect of display and environmental illumination on bottom up visual attention.

To put these areas in context Figure 4.2 shows the selective rendering pipeline, with the areas this thesis seeks to address outlined in green.

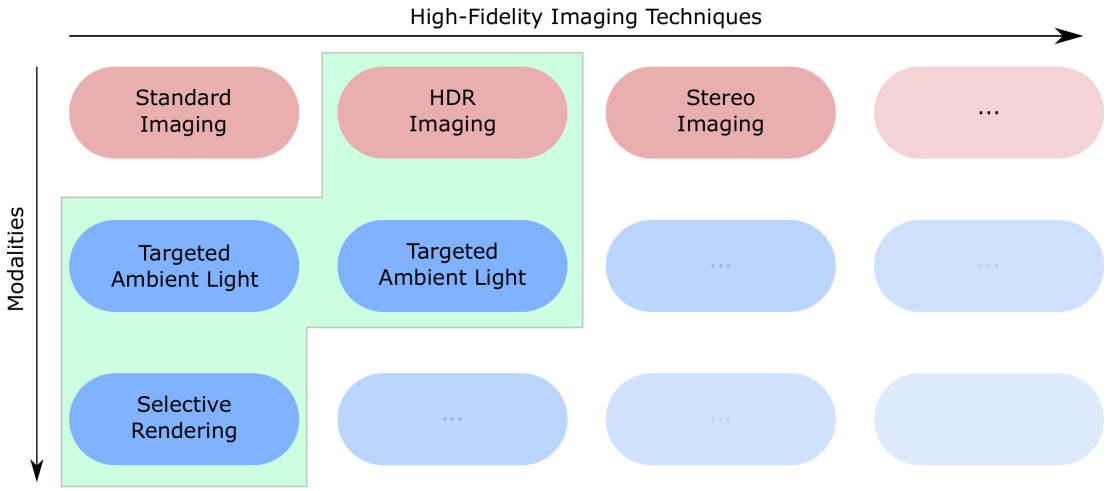


Figure 4.2: The intersection of visual attention and high-fidelity imaging

4.3.1 Research Question

The research question that this thesis seeks to answer is:

- How can in-depth knowledge of visual attention be used effectively when generating high-fidelity imagery?

To answer this there are three objectives:

- Develop and test a method for selective rendering utilizing Bidirectional Reflectance Distribution Functions.
- Create an HDR eye tracking data set with multiple uniform enforced brightnesses.
- Analyse the effect of HDR and ambient light on visual attention and visual fatigue.

4.4 Thesis structure

In order to accomplish the objectives laid out in the previous section, three experiments were conducted. Their methods and content are explained in the following three chapters.

In order to address the first research question, Chapter 5 describes a method for adaptive selective rendering using saliency maps to alter the quality of BRDFs used

in rendering a scene, this was implemented in a physically accurate ray tracer and the validated with a psychophysical experiment.

Chapter 6 follows, detailing an experiment to evaluate the effect of high peak screen brightness levels on fixation patterns of observers. This involves two stages, the collection of an HDR eye tracking dataset and the analysis of the recovered fixation maps. This addresses the second research question.

To answer the third research question, Chapter 7 examines the effect of higher brightness and environmental illumination. This is to further aid in the targeting of saliency models for different rendering applications.

Chapter 5

Selective BRDFs for High Fidelity Rendering

High fidelity rendering systems rely on accurate material representations to produce a realistic visual appearance. However, these accurate models can be slow to evaluate. This chapter presents an approach for approximating these high accuracy reflectance models with faster, less complicated functions in regions of an image which possess low visual importance. A subjective rating experiment was conducted in which thirty participants were asked to assess the similarity of scenes rendered with low quality reflectance models, a high quality data-driven model and saliency based hybrids of those images. In two out of the three scenes that were evaluated significant differences were not found between the hybrid and reference images. This implies that in less visually salient regions of an image computational gains can be achieved by approximating computationally expensive materials with simpler analytic models.

5.1 Introduction

There is a continuing demand for increased accuracy and simulation speed in virtual environments. This comes at the cost of ever increasing demands on computational resources, especially when using algorithms relying on ray-traced lighting and in scenes with multiple complicated materials. These demands can be partially mitigated through

various strategies, such as improving light transport algorithms, coherence for tracing and shading, filtering, and reducing computation in less visually important regions of the scene. This chapter investigates the latter approach, specifically whether straightforward and computationally inexpensive surface reflection models can be used in place of accurate but more detailed models in less salient regions of the image.

Visual attention models provide a framework to predict the areas of an image which are likely to be attended to by the Human Visual System (HVS). These models have been applied to improving rendering performance by several previous authors [CCW03][CDMPdS07][KDCM14][GDS14] as discussed in Section 2.2.4, however this work has predominantly focused on decreasing the number of samples computed in a Monte Carlo image synthesis context. Our work aims to improve performance in an orthogonal manner; by reducing the complexity of materials in less visually salient areas, computational savings can be gained without a significant loss in perceived quality of the resultant image.

Material models used in physically-based rendering are known as *Bidirectional Reflectance Distribution Functions* (BRDFs). These are 4D descriptions of how incident light is reflected in a given direction. Two types of models are frequently used in rendering systems; analytical and data-driven. Analytical BRDFs range from simple models of diffuse reflectance, to parametrised glossy models, such as Phong [Pho75] and Walter [WMLT07], and multi-lobed BRDFs [LFTG97]. These parameters can be altered to affect the appearance of multiple surfaces, but can only approximate real materials. Data-driven BRDFs consist of captured reflectance values of real-world materials. These typically consist of a 4D (or 3D in the case of isotropic materials) lookup table which is queried at runtime. Uncompressed data-driven BRDFs require significant storage, for example the MERL database [MPBM03] stores each measured surface as a 33MB binary file, the total size of the database of 100 materials being 3.30 GB. This requires significant memory bandwidth during the frequent lookups from the table.

Typical rendering systems additionally use mixtures of these BRDFs, often in a spatially varying [Mca02], layered [WW07], or as a mixture driven by a shader. Evaluating shading on a surface can take longer than tracing rays [ENSB13], and

therefore savings in BRDF evaluation can potentially significantly speed up the rendering process. The focus of this chapter is on approximating data-driven BRDFs with analytical models in less salient regions. We choose data-driven BRDFs as a baseline as they are expensive to evaluate and provide accurate measured data.

In this chapter, an application of visual saliency models to reduce the computational and memory bandwidth costs for BRDF evaluation is described. A series of analytic BRDF models were fitted to the MERL dataset, and applied to several virtual scenes. A saliency map was generated for each scene and used to evaluate whether users notice a difference between an expensive, but accurate, data-driven BRDF, and a fitted analytic approximation.

The structure of this chapter is as follows: Section 5.2 outlines the methodology that was used to prepare the scenes and reflectance models, as well as the procedure and design of the perceptual experiment. The results of the perceptual tests are given in Section 5.3, discussed in Section 5.4, and conclusions are drawn in Section 5.5.

5.2 Methodology

This work is motivated by the need to provide better overall performance for scenes that may not require detailed material representations at every point in the scenario. In order to demonstrate the feasibility of such a system an experiment was conducted to identify perceptual differences amongst analytical and data-driven BRDFs and hybrids of the two that use the higher quality BRDF in areas of the region considered more salient.

5.2.1 Design

The experiment is a subjective rating study, in which participants were asked to rate the quality of images in comparison to a ground truth image on a scale of 1 to 100. A hidden reference is also included to provide a relation to the ground truth image. The rating design permits quantification of distance between methods. The hidden reference enables comparisons with the other stimuli to identify perceivable differences across them.

The scale used in this experiment offers a sufficient breadth for participants to give a wide range of ratings. Each participants' scores are normalised when calculating results in order for effective comparisons to be made.

The independent variables are the analytical BRDF which is used to create the mixed images and the scenes used. Both independent variables follow a within-participant design. The BRDF variable consists of seven possibilities, three analytical BRDFs, a Diffuse model (D), the Phong BRDF (P) [Pho75] and the Walter BRDF (W) [WMLT07], three mixed Saliency models consisting of the three chosen BRDFs (SD, SP, SW respectively) mixed with the data-driven BRDF using a saliency map and a hidden Reference (R), see Figure 5.1. The three scenes used were of a kitchen, a conference room, and a lounge, see Figure 5.2.

The camera angle, BRDF fitting procedure, sampling algorithm, saliency model, image resolution, number of samples per pixel and viewing time for each image were set as constant across all scenes and reflectance models. The dependent variable is the rating given to each stimulus.

5.2.2 Materials

This section describes the preparation of the materials, particularly the stimuli used in the experiment.

BRDFs

The choice of the data driven BRDF used in these experiments is primarily motivated by three factors, the number of available materials, the density of the measurements and the focus of related literature. Therefore the MERL database was chosen as the reference BRDF. The Walter BRDF was chosen because it is heavily used in industry, fits the MERL database well and is a good representation of micro-facet BRDFs. The Phong BRDF was also chosen for its prevalence in rendering and its speed. It is purely specular and is normally combined with a basic lambertian diffuse model to provide colour.

For the purpose of this experiment a densely populated data driven BRDF (the

MERL database) was chosen as a reference, as using an analytic model would require an extra level of fitting. For this purpose Löw's [LKYU12] fitting function is adapted to fit to the methods that are used in the perceptual experiment. It was chosen for its ability to fit to simple models rather than to the Approximate Bayesian Computation (ABC) method for which it was designed.

(a) Diffuse (D)

(b) Phong (P)

(c) Walter (W)

(d) Salient-Diffuse (SD)

(e) Salient-Phong (SP)

(f) Salient-Walter (SW)

Figure 5.1: The Kitchen scene, rendered with the methods used in the experiment. The zoomed area shows the dragon sculpture, represented with the chrome MERL material

This section outlines the procedure that was followed to fit the analytical BRDFs

used in this study to each entry in the MERL database. Each model had independent parameters for each colour channel, with the exception of roughness parameters, as surface micro-structure is independent of the incident wavelength. In the case of the Walter BRDF, the incident index of refraction was the same for each colour channel as the incident medium is air and the difference in refractive index for red, green and blue's respective wavelengths is negligible.

The MERL BRDFs have an angular resolution of one degree, as the materials are isotropic, the reflectance doesn't vary with the incident angle, ϕ, therefore $\phi = -\phi$. The density of the tabulated BRDFs is then $90 \times 90 \times 180 = 1,458,000$ entries for each colour channel.

In order to find parameters for the Diffuse, Phong and Walter BRDFs a nonlinear least-squares regression algorithm is used of the form:

$$\min_{p \in R^n} g(x, p), l < p < u \tag{5.1}$$

where x represents the fixed inputs to the function and p, l and u are vectors of length m containing the parameters for the BRDF in question and the lower and upper bounds of those parameters, respectively. The cost function, g, is defined as:

$$g(\omega_i, p_k) = \sum_{i=1}^{n} E_i^2 \tag{5.2}$$

where E_i is the squared error at the i^{th}data point on the k^{th} iteration and ω_i is the vector representing the direction of incoming light. To calculate the error a variant of the formulae described in the paper of Löw et al. is used.

$$E = \sin\theta_o \sqrt{y_m^2 - y_a^2} \tag{5.3}$$

where θ_o is the elevation angle of the view vector ω_o, expressed in spherical coordinates and y_m and y_a are the weighted outputs of the material and the analytical BRDF being fitted, respectively.

$$y_m = \ln(1 + \cos\theta_i f_r(\omega_i, \omega_o)) \tag{5.4}$$

$$y_a = \ln(1 + \cos\theta_i f_r(\omega_i, \omega_o, p)) \tag{5.5}$$

where θ_i is the elevation angle of the lighting vector ω_i and p is a vector containing the parameters for the analytical BRDF to be fitted.

In this weighting function the MERL material and analytical BRDF are queried and their returned RGB values are multiplied by the cosine of the incident vector's elevation, to reduce the effect of the poor data near grazing angles [NDM05]. Then this result is logged, in order to put an even weight on specular and non specular regions. The squared difference of each colour channel is then calculated.

In order to gather a manageable set of vectors to sample the material circles of increasing radii are projected from the unit disk onto the hemisphere centred around the perfect specular direction. Half of the projected circles have a small radius (approximately 0.2 on the unit disk), in order to capture the specular colour and glossiness and half have a larger radius, to capture the diffuse colour.

Scenes

The scenes chosen were all enclosed, indoor scenes. The three scenes were of a kitchen, to the best of the author's knowledge this is a model reproduced from the original by Wang et al [WWZ+09], a conference room, originally created by Greg Ward and stored in the Maguire repository [McG17], and a lounge, the model for which was created by the author's research group. The kitchen, conference room and lounge scenes contained 68, 26 and 54 distinct materials, respectively. The reference images for each scene can be seen in Figure 5.2 and all images for the Kitchen scene can be seen in Figure 5.1 .

Salient Mixture Model

The mixture model stimuli are produced as hybrids of the MERL materials and analytical BRDFs. The saliency model employed is *Graph-Based Visual Saliency* (GBVS) [HKP06]. However this method is agnostic of the method and any other image-based visual saliency model could be used. An example saliency map, of the Kitchen scene, is shown in Figure 5.3. The mean saliency for the Conference, Kitchen and Lounge scenes was 0.1918, 0.2097 and 0.2165, respectively, where a value of one would represent a pixel with 100% salieny and a value of zero would represent a 0% saliency. This means that

(a) Conference

(b) Kitchen

(c) Lounge

Figure 5.2: Reference images, materials rendered using samples from the MERL database

on average the reference method was sampled 19.18%, 20.97% and 21.65% of the time for the Conference, Kitchen and Lounge scenes respectively.

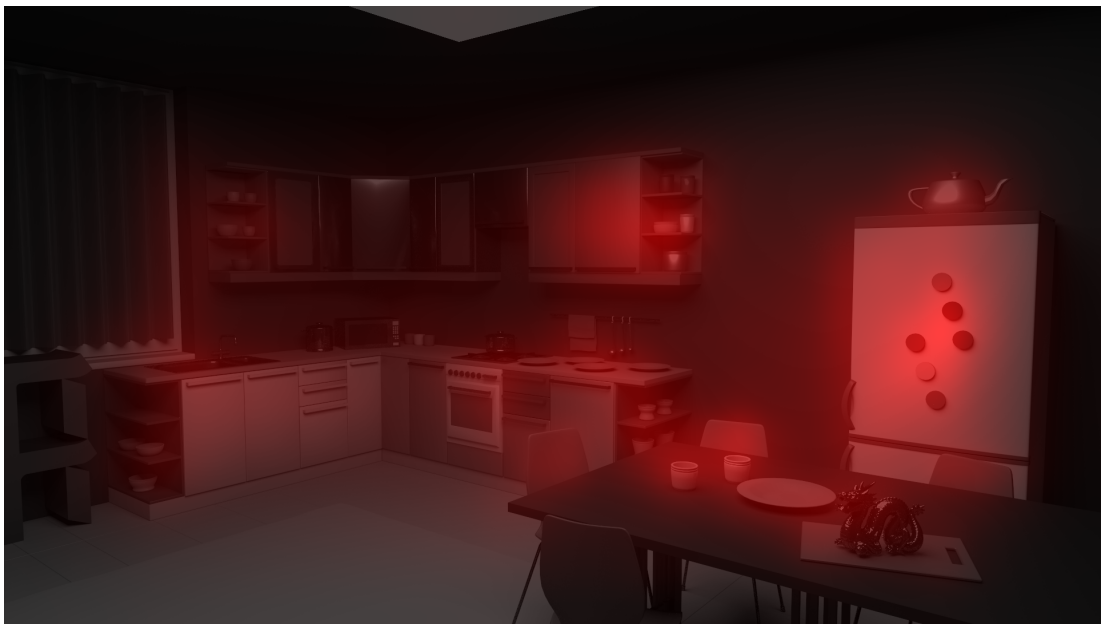


Figure 5.3: The Kitchen scene with the saliency of each pixel overlayed in red. Brighter red values represent a higher saliency value at that pixel.

The salient mixture images are created by using the saliency of the given pixel to weight the BRDF that is sampled. Every time a surface is to be shaded, either the data-driven BRDF or an analytic model is selected based on sampling the normalised saliency value:

$$fr = \begin{cases} fr_{DATA} & \text{if } \xi < S(i) \\ fr_{ANALYTIC} & \text{otherwise} \end{cases} \tag{5.6}$$

where fr is the selected BRDF, fr_{DATA} is the relevent data driven BRDF, $fr_{ANALYTIC}$ is the analytic BRDF (Diffuse, Phong or Walter), $S(i)$ is the saliency normalized to [0,1] at pixel i, where the path originated, and ξ is a uniformly distributed random number between 0 and 1.

5.2.3 Participants

There were 30 total participants in this experiment, 4 female and 26 male from a variety of academic backgrounds. Only one possessed expert-level graphics knowledge. All participants had normal or corrected to normal vision. The participants were offered no incentive to take part in the study.

5.2.4 Procedure

The experiment took take place in an office environment with low ambient light varying between 8.1 and 35.3 lux and the images viewed on a 55 inch HD monitor with the brightness and contrast set low to avoid eye fatigue, as per the ITU-R recommendation [Ass03]. The viewing distance was 2.1 metres to avoid discomfort, as per the same recommendations.

A sequential arrangement for the images was followed with the ground truth image displayed first followed by the stimulus. Both were displayed for a duration of three seconds [MTM12]. All stimuli were shown to all the participants.

This was followed by a five second break for the participants to rate the image before the next image in the series was displayed. The order of the images was randomised. Each experiment took around ten minutes to complete.

5.3 Results

In this section the results of the perceptual experiment are presented, analysed and discussed; as timings for rendering both the hybrid images and the images rendered with each method.

5.3.1 Timing

In order to assess the efficiency of each method, the time trials were run on a single CPU core. Timing results were conducted on a single thread using an Intel Xeon E5-2620 at 2.00GHz with 16gb of RAM on the Windows 7 64-bit operating system. The images were rendered at a resolution of 1920×1080 with one direct and one indirect lighting

Table 5.1: Time taken (s) for one sample per pixel in a 1920×1080 image for each BRDF in each scene. Below are speedups of time taken to render with the reference method compared to the relevant method.

	Scene	D	P	W	SD	SP	SW	R
Timings (s)	Conference	16.27	16.95	18.08	16.69	17.20	17.71	18.63
	Kitchen	13.11	13.96	15.00	13.8	14.20	14.69	15.62
	Lounge	21.52	21.87	22.72	21.74	22.30	22.71	23.41
Speedup vs R	Conference	1.146	1.099	1.031	1.116	1.083	1.052	1.000
	Kitchen	1.191	1.119	1.041	1.131	1.100	1.063	1.000
	Lounge	1.088	1.070	1.030	1.076	1.050	1.030	1.000

bounce. The timings for computing one sample per pixel are shown in Table 5.1. The images were rendered for 100 samples per pixel and the timings were then averaged to give the results below. The ratio of computation time vs. the reference is shown in the lower half of Table 5.1. This illustrates that computational gains can be expected using the saliency weighted methods.

5.3.2 Perceptual Tests

In order to analyse the similarity of images to the hidden reference the ratings given by participants are converted into distances between their rating of each condition and their rating of the relevant reference image, as recommended by Mantiuk et al. [MTM12].

$$d_{i,j,k} = r_{i,ref(k),k} - r_{i,j,k} \tag{5.7}$$

where d is a distance score, indicating the distance between a given participant's raw rating, r, of an image and their rating of the respective reference. Here i, j and k represent the observer, image and scene respectively.

The scale used in this study, 1 to 100, is broad and different participants have different standards for what a low similarity score is. Therefore in order to more effectively compare results from different participants, Z-scores are calculated, see equation 5.8.

Table 5.2: Mean Z-scores for each scene.

Scene	D	P	W	SD	SP	SW	R
Conference	0.670	0.404	0.281	0.009	0.319	0.270	-0.550
Kitchen	1.121	-0.127	-0.012	-0.328	-0.288	-0.440	-0.550
Lounge	1.039	-0.047	0.231	-0.590	-0.501	-0.364	-0.550
Overall	0.943	0.077	0.167	-0.303	-0.157	-0.178	-0.550

Table 5.3: Contrast comparisons between BRDFs in each scene. Coloured groupings indicate no significant differences across methods. *significant at $p < 0.01$

Scene	**BRDF**	**Kendall(W)**
Conference	*D P W SP SW SD R*	0.213*
Kitchen	*D W P SP SD SW R*	0.227*
Lounge	*D W P SW SP SD R*	0.267*
All	*D W P SP SW SD R*	0.397*

This gives a measure of the extent to which participants thought a given image differed from the relevant reference image and sets the mean and standard deviation of each participant's ratings across all scenes to 0 and 1, respectively.

$$z_{i,j,k} = \frac{d_{i,j,k} - (\bar{d}_i)}{\sigma_i} \tag{5.8}$$

where σ_i is the standard deviation of the i^{th} observer's ratings across all images.

Figure 5.4 displays the mean scores for each method across all scenes. An average negative z-score reveals that a method was deemed to be above average by the participants, where zero is the average rating. The scores for the hidden reference are also included as participants could have rated the other images as more similar to the reference than itself. The reference images have scores of $\frac{-\bar{d}_i}{\sigma}$ for the i^{th} participant, by definition.

Results were analysed via two-way repeated measures ANalysis Of VAriance (ANOVA)

in a 7 (method) $\times$ 3 (scenes) factorial design. The main effect of scenes did not violate the assumption of sphericity (Mauchly's Test of Sphericity, $p > 0.05$). The main effect of the scene did not produce significant differences, $F(2, 384) = 3.018$, $p > 0.05$, indicating no significant differences were found between scenes.

The main effect of method did not violate the assumption of sphericity (Mauchly's Test of Sphericity, $P > 0.05$) and was significant $F(6, 384) = 29.445$, $p < 0.01$. This suggests that the method used had a significant effect on the participants' ratings.

Kendall's co-efficient of agreement [KS39] was computed on the three scenes and on the collapsed overall scores to identify agreement across participants. Kendall's co-efficient gives a value of 0 when participants are in complete disagreement and 1 when in complete agreement. The results are shown in Table 5.3. The results are all considered significant ($p < 0.01$) indicating relative agreement in judging by the participants.

Pairwise comparisons with *Bonferroni corrections* were conducted to identify significant differences amongst the individual methods.

The results of these tests can be seen in Table 5.3. Coloured groupings represent methods between which a significant difference was not found. Results demonstrate groupings of the saliency methods and the other methods. As expected Diffuse by itself performs poorer than all other methods, the Phong and Walter methods are grouped together as are all the mixed model methods. These groupings represent the lack of a discovered significant difference between how similar images created using different methods were to the reference image. Finally the Reference scored better than the rest of the methods, though this difference was not found to be significant in some of the scenes, as seen in Table 5.3.

5.4 Discussion

The aggregate scores of all participants for each method are displayed in Figure 5.4. Here large negative results indicate a higher rating was given to the method. The pink bars on the right hand side of the figures represent the average ratings of the reference methods, for the relevant scene. The salient methods are significantly better than their

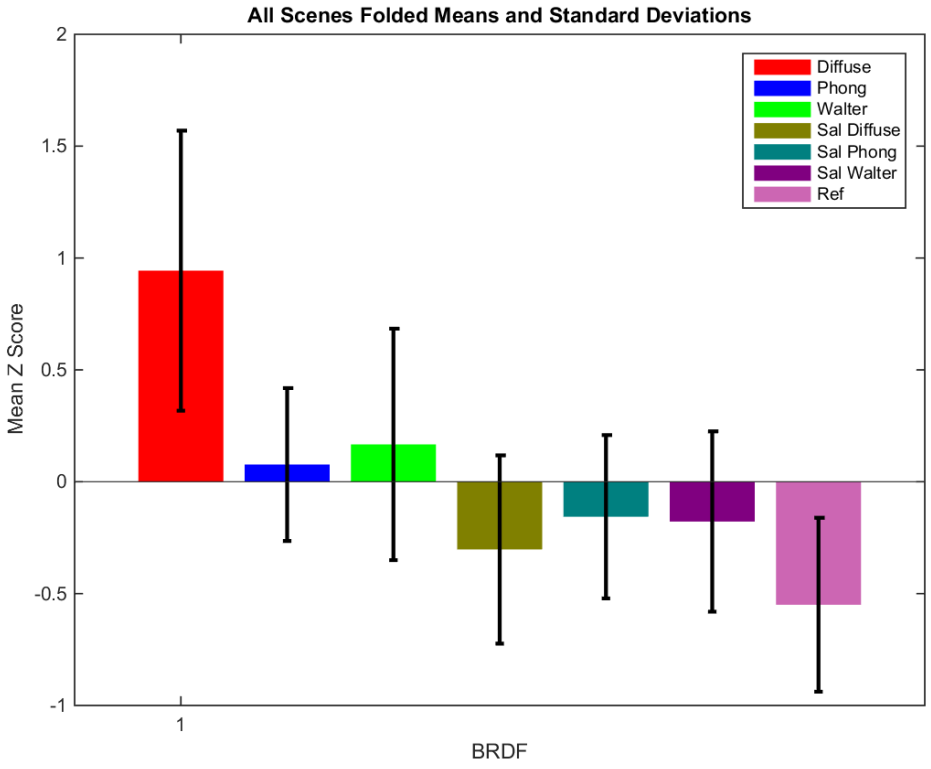


Figure 5.4: Aggregate scores and standard deviation for each BRDF across all participants and scenes

counterparts, overall, but there was no significant difference to discriminate amongst them with the current set of captured data.

An examination of Table 5.2 shows the results for the Conference scene differ heavily from the Kitchen and Lounge scenes. In the conference scene all methods were rated lower than the reference and pairwise comparisons found a significant difference in the similarity ratings of the reference image and the hybrid images and between the hybrid images and their analytical counterparts, as can be seen in Table 5.3. In this experiment the saliency based methods did not exhibit significant difference in every scene and the overall average, suggesting they are of a similar quality. In addition Phong and Walter exhibited no significant difference while Diffuse has no correlation with any other method except in the conference scene.

The timings in Table 5.1 show that a decrease in computation cost can be expected across all methods. As can be seen in Table 5.2 the salient diffuse method achieved the greatest speed ups of the hybrid methods with savings of 7.6%, 11.6% and 13.1% for an average speed up of 10.8%. The Phong and Walter methods also outperformed the reference with average speed ups of 7.8% and 4.8% respectively. As Table 5.3 shows, visual perception of the scene is not significantly affected.

5.5 Conclusion and Future Work

This work has investigated whether analytical BRDFs can be substituted for data-driven models in a rendering pipeline without a noticeable effect to an observer. Three scenes and seven methods for representing materials were compared through the utilisation of a subjective study and statistical analysis, in two of the three scenes examined, significant differences were not found between the reference image and in all scenes the hybrids outperformed their non-salient counterparts. All three salient methods provided an improvement in computational performance with salient diffuse providing the largest mean computational saving (10.8%). This provides an indication that visual saliency can be used to improve computational performance by replacing expensive materials with cheaper analytic models in less visually important regions of the image.

This work could be expanded to approximate high quality BRDFs that incorporate

physical approximations of diffraction [LKYU12][HP15] and the polarisation of incident light [HTSG91]. These functions are more mathematically complicated and in some cases [HP15] require precomputation of the geometric shadowing and masking functions.

Chapter 6

HDR Eye-tracking

While visual attention has applications in many areas in film, video games, photography, advertising and the military, the central concept of selective rendering is that visual attention models can be used to reduce the cost of rendering a scene, without a significant drop in perceived quality. This is reliant on the model being used being an effective estimator of attention.

While recent eye-tracking studies [BPT10, DNPN14, NKHE15] have evaluated the deficits that existing models of visual attention exhibit when applied to High Dynamic Range (HDR) imaging, these studies fail to take into account the effect of varying the peak luminance at which the image is displayed. HDR images are scene referred, which means they are intended to represent the real world luminance of the scene they are recreating. Thus, the studies mentioned previously have considered HDR a binary variable, rather than evaluating the effects of fixed HDR luminance ranges on visual attention. In this work an eye tracking study was conducted across discrete luminance values from 500cd/m^2 to $4,000\text{cd/m}^2$. Participants viewed a sequence of HDR images at varying peak luminance levels on an HDR display, while their eye movements were recorded with an eye-tracking device. This chapter presents the results of this study and discusses the interdependence of display brightness and low level saliency.

Results of a statistical analysis of the eye tracking data found evidence of a significant difference in viewing strategies of LDR and HDR content, with the reliability of an LDR baseline falling with an increasing screen luminance. However, the results also suggested

that the reliability of eye tracking data also decreases as display luminance increases.

6.1 Outline

The advent of High Dynamic Range imaging has brought with it a number of visual attention models and eye-tracking studies [BPT10, DNPN14, NKHE15, DPN16, BDDPN15], however these studies fall short of experimentally determining if there is a need for HDR eye-tracking data.

This chapter presents an HDR eye-tracking dataset, in which participants viewed images at four distinct brightness levels, and presents a statistical analysis of the effect of peak screen luminance on visual attention. These studies focus on establishing if existing visual attention models are applicable to HDR content, as HDR images store luminance in a fundamentally different manner to LDR images, as discussed in Chapter 2, it is necessary to adjust the ways in which existing LDR visual attention handle luminance. While useful, an analysis of visual attention models from an image processing perspective, as in these studies, does not establish whether the fundamental mechanisms of visual attention are affected by HDR luminance levels. This work addresses this in two primary manners, first through a thorough analysis of the impact of targeting HDR at specific luminance levels, and second by providing a dataset of fixation maps, at varying luminance levels, to aid in the verification of future models.

6.2 Methodology

This section outlines the design, materials and procedure for the creation of an eye-tracking dataset, which will be used to test two hypotheses; does the power of LDR fixation maps to predict new fixations change as the brightness of the stimulus increases; if there is a difference, are fixations from HDR datasets better predictors of HDR fixations than their LDR counterparts, which are commonly used for saliency map validation. The dataset consists of 80 images from the Fairchild database [Fai07] at four different brightness levels, from 500 to 4,000 $\mathrm{cd/m^2}$. All forty eight participants viewed all images, split evenly and randomly across each of the four conditions, such

that 20 images were shown for each condition per participant. The aim of the collection is to provide the necessary data to examine the effect of screen brightness on visual attention, which will occur in the next section.

6.2.1 Design

Hypothesis 1 (H1): *The brightness of a display will have a significant effect on the fixation patterns of observers.*

Hypothesis 2 (H2): *Eye-tracking data captured for HDR content will provide a more accurate baseline for predicting new fixations on that content than a standard LDR baseline.*

In order to assess the effect of luminance on visual attention an eye-tracking experiment was conducted. Participants were asked to view a sample of images at different brightness levels while having their eye movements recorded by an eye-tracker. The independent variable, peak display *brightness* of the viewed image, had four levels: 500, 1,000, 2,000 and 4,000 $\mathrm{cd/m^2}$. The first level, 500 $\mathrm{cd/m^2}$, was chosen to be in line with the peak output of common LDR displays, the last, 4,000 $\mathrm{cd/m^2}$, for the peak output of the available HDR monitor. These were distributed in a within-participants design to observe effects on the dependent variable, the viewing strategies of observers, which is measured by a comparison of their *fixation maps*.

The sample consisted of eighty images, which were randomly divided into four groups of twenty, one at each of the four levels, for each participant. The participants viewed all twenty images for a condition before moving onto the next. Every possible permutation of the four conditions was shown to at least two participants, and the permutation was chosen randomly from the remaining permutations for each participant. The distance to the screen (1.8m), the display duration of each image (10s), the instructions given to participants, the ambient illumination level (0lx), and the display used were kept constant. The order in which images were displayed was also randomised. The (10s) display time was chosen to minimise the length of the viewing time, to help protect against eye-strain, and boredom. In addition this viewing period is in line with the fixation density map comparison carried out by Engelke et al. [ELW+13].

The experiment was conducted in a controlled laboratory environment. The ambient illumination in the room was measured as 0lx.

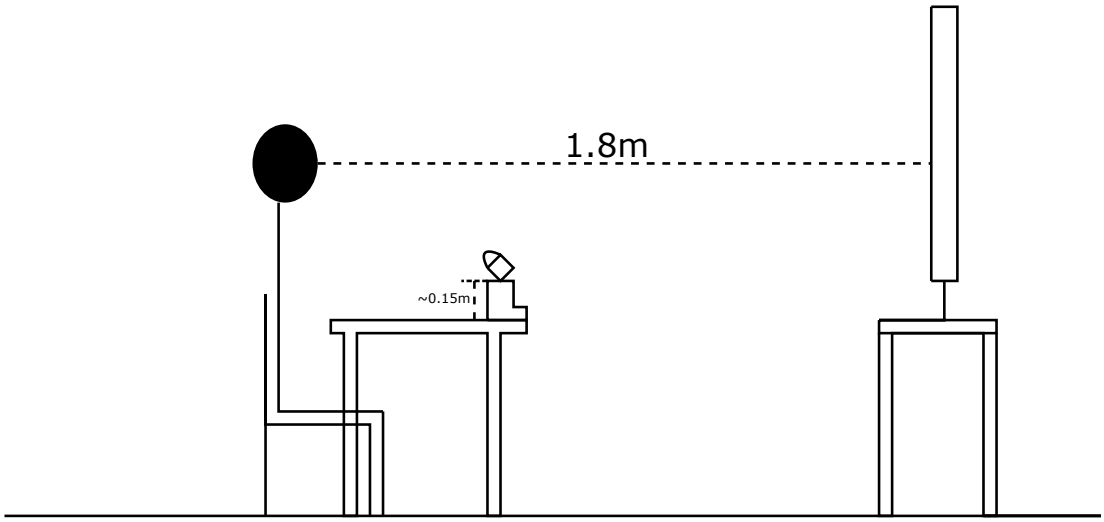


Figure 6.1: Layout of laboratory environment

6.2.2 Materials

The eye-tracking data-collection was conducted in a dark room, with low ambient luminance (0lx). The images were displayed on a SIM2 HDR display [SIM15] with a peak luminance of 4,500 $\mathrm{cd/m^2}$ resolution of 1920×1080 and an active screen area of $1021mm \times 572mm$.

The image sample consisted of eighty HDR images taken from the Fairchild database [Fai07]. Each image was scaled so that the 99^{th} percentile of brightness for image matched each of the four brightness levels. Pixels with brightnesses above this level were clamped. This has the potential to desaturate the image, however it guarantees a uniform peak brightness across images.

The eye tracking software used in this study was the Seeingmachines Facelab 4.5 package. This software was used in conjunction with two Flea cameras [Poi17], with IR filters attached, and an IR light pod mounted in the centre of the cameras. See Figure 6.2 for a top-down view of the scene.

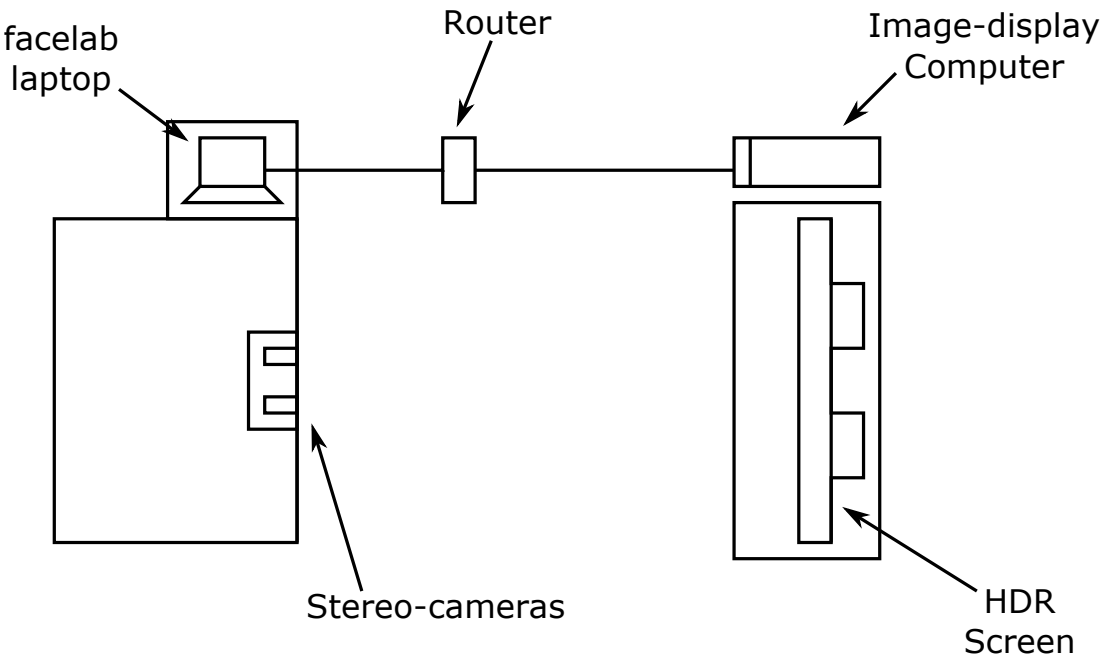


Figure 6.2: Top-down view of experiment setup

6.2.3 Procedure

Upon entering the test environment participants were given time to review the participant information sheet, then asked to sign a consent form and sit in a chair in front of the HDR display for two minutes as a model of their head was created using a stereo-camera pair. Once this model was created, the participants were asked to focus on nine points on the screen, to calibrate their gaze.

Before viewing the image sequence the participants were read the following script: "You are being asked to view a sixteen minute slide show, after which you will be asked to answer a few general questions about what you have seen."

The participants were then asked to view the image sequence for a single permutation of conditions. Each image was displayed for ten seconds, with a two second gap in between images to avoid contamination across images. During the 16 minute duration of the experiment the eye-tracker recorded the participant's eye movements.

6.2.4 Participants

Forty-eight university students, staff, employees, and members of the public participated in the experiment. All participants participated voluntarily, at the request of the writer. Eighteen of the participants were female, thirty were male. All the participants had normal or corrected to normal vision. The average reported age of all the participants was 28.0 years.

6.3 Results

This section details the results of the eye-tracking experiment, seeking to address the hypotheses stated in previous sections. Namely, is there a need for HDR tracking datasets and if so, to what extent?

To answer these questions, first the nature and organisation of the data to be analysed is explored in Section 6.3.1. Following this a statistical analysis is presented contrasting viewing strategies at various *brightnesses* with an LDR-baseline using a number of pairwise metrics, discussed briefly below and covered in greater depth in Section 3.4, saliency map evaluation. After which, k-fold cross validation is used to assess the uniformity of observers attention at varying brightness levels.

Finally the results of the k-fold cross-validation is contrasted with the results of the LDR-baseline analysis to evaluate the potential benefit of using HDR data instead of traditional LDR datasets for evaluating the accuracy of visual attention models.

6.3.1 Processing fixation data for analysis

There are two forms fixations take during analysis, the first is a fixation map, this is a binary map, with the same resolution as the viewed image. Pixels that were attended have a value of one, and pixels that were not attended have a value of zero. These are created using Algorithm 1, informed by the work of Engelke et al. [EMZ09]. The second, and more common format, is a fixation density map. These are similar to fixation maps, in that they represent the locations in an image that a participant focused on, however account for the visual angle of the fovea, which is around 2° [NKHE15, EMZ09, JDT12],

by blurring the fixation map with a gaussian filter with a sigma value of half the visual angle. The 2° visual angle was calculated to be approximately 118 pixels for the apparatus configuration in this experiment.

Fixation maps for individual participants are gathered by condition for each image before being convolved with the aforementioned gaussian filter to create a single fixation density map for each condition. This process is illustrated visually in Figure 6.3, 6.4, and 6.5. In the interest of clear presentation, only eight participants' fixation data is shown per figure as opposed to the full twelve per condition.

Algorithm 1: Pseudocode for creation of fixation maps

Define cluster threshold τ

Define minimum number of fixations F_{min}

set $j = 0$

Initialise first cluster C_j

for $i = 1$ **to** N_{gp} **do**
    $\mu_j = mean(gp_i + \mu_j)$
    if $\delta_i < \tau$ **then**
        $C_j \leftarrow gp_i$
    else
        $j = j + 1$
        $C_j \leftarrow gp_i$
    end
end

for $k = 1$ **to** $length(C)$ **do**
    n_{gpc} = number of GP in C_k
    if $n_{gpc} < F_{min}$ **then**
        $F_n = C_k$
    end
end

In order to evaluate the collected fixation maps statistically, quantitative metrics are required. Six common metrics were chosen, these are the Area Under the receiver operating characteristic Curve (AUC), specifically the variant from Judd et al. [JEDT09],

a thresholded variant of the Earth Mover's Distance (EMD) [RTG00] [PW09], the Similarity metric [SB91], Pearson's Correlation Coefficient (CC), Normalised Scanpath Saliency (NSS) [PIIK05], and a symmetric form of Kullback Leibler Divergence (KLD).

The metrics used for fixation pattern analysis are pairwise metrics, that is they compare either two probability distributions (EMD, KLD, Pearson's Correlation Coefficient and Similarity) or test how well a set of fixations fit a saliency map (AUC, NSS), or FDM. Each method gives some measure of the fitness or correlation of two sets of fixations. Therefore, extracting quantitative data from them requires comparing fixations from two conditions, or two subsets of a given condition.

6.3.2 Inter-Brightness Comparison

To examine the effect of brightness on the reliability of LDR eye tracking datasets for HDR content, the LDR 500 $\mathrm{cd/m^2}$ condition from each image was taken as a baseline for the pairwise comparisons. Therefore the resulting scores represent each ratings measure of the similarity of the other conditions to condition A, Table 6.2 illustrates this visually. This provides a measure of how similar the fixations for each condition are to the fixations for the LDR data.

The analysis was split into two categories, *initial fixation* and *free viewing*, with the former consisting of only the first two seconds of display and the latter the full ten second display period. These are presented concurrently but not contrasted.

The fixation density maps for each brightness of ten scenes can be seen in Table 6.1.

The results of a repeated measures ANOVA of the six metrics' scores can be seen in Table 6.3. The data didn't violate the assumption of sphericity, with the exception of the CC results, for which a Greenhouse-Geisser adjusted score is reported, and marked with an asterisk. The F-score in this case was recorded as $F(1.78, 140.67)$. The results of the Bonferroni-corrected pairwise comparisons can be seen in Table 6.4, where conditions are arranged from left to right in order of decreasing mean similarity to the LDR baseline. Figure 6.7, 6.8, 6.9, 6.10, 6.11 and 6.12, display the mean scores for each condition across all images visually for comparison. The initial fixation period is labelled as "Full 2s" and the full ten second free viewing period is labelled as "Full 10s". As can be seen

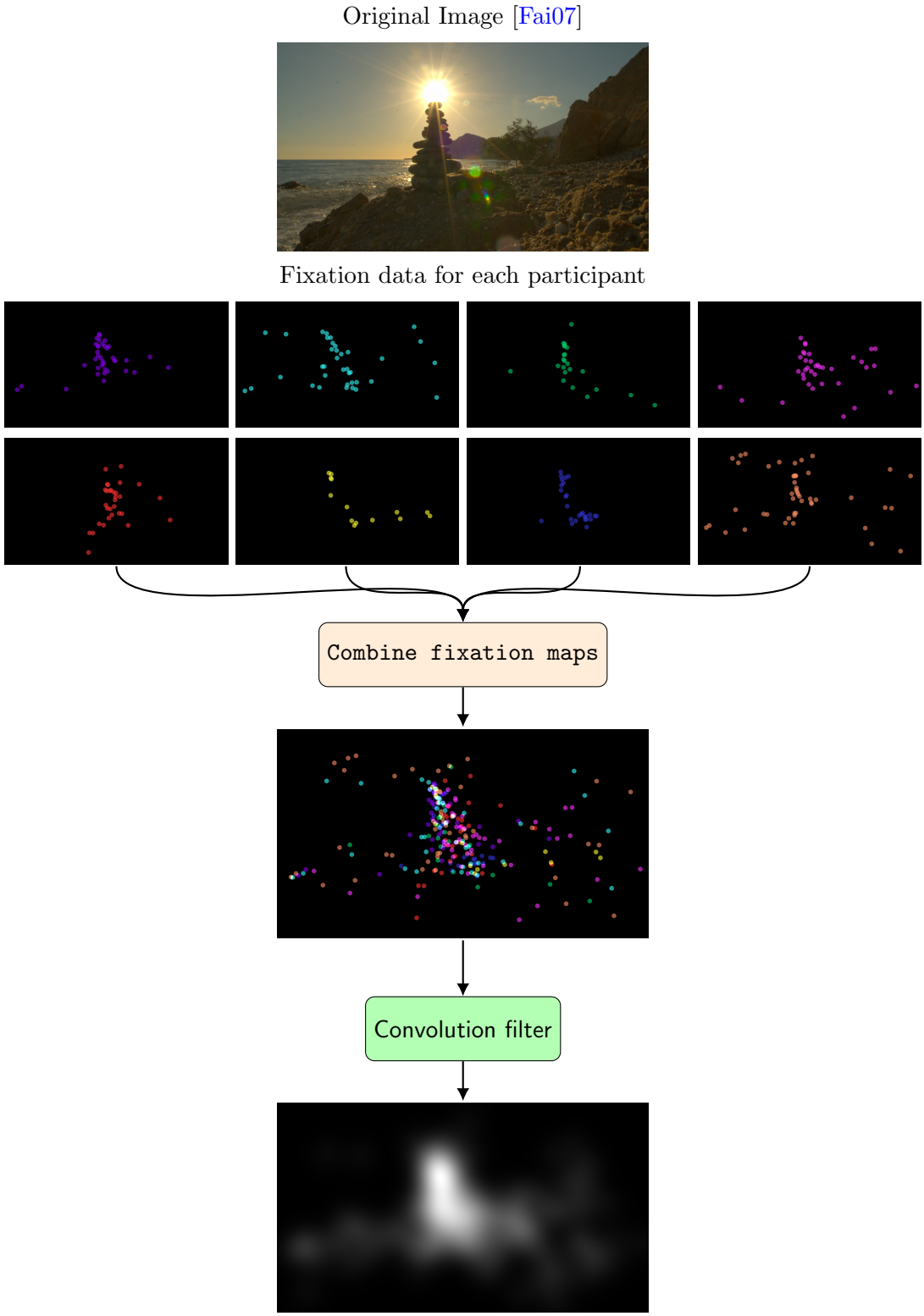


Figure 6.3: An illustration of the fixation map creation process for the StoneTower [Fai07] image

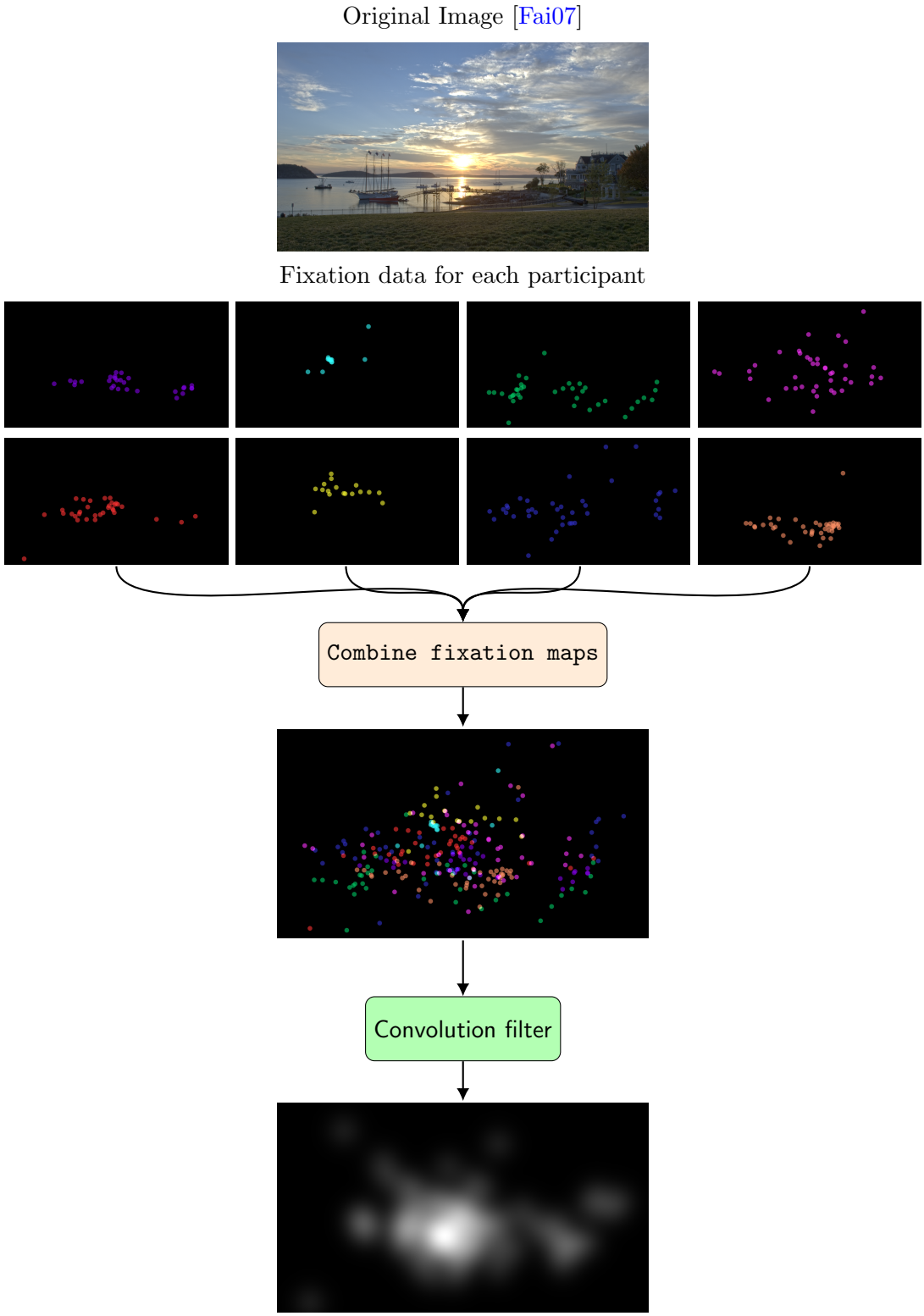


Figure 6.4: An illustration of the fixation map creation process for the BarHarborSunrise [Fai07] image

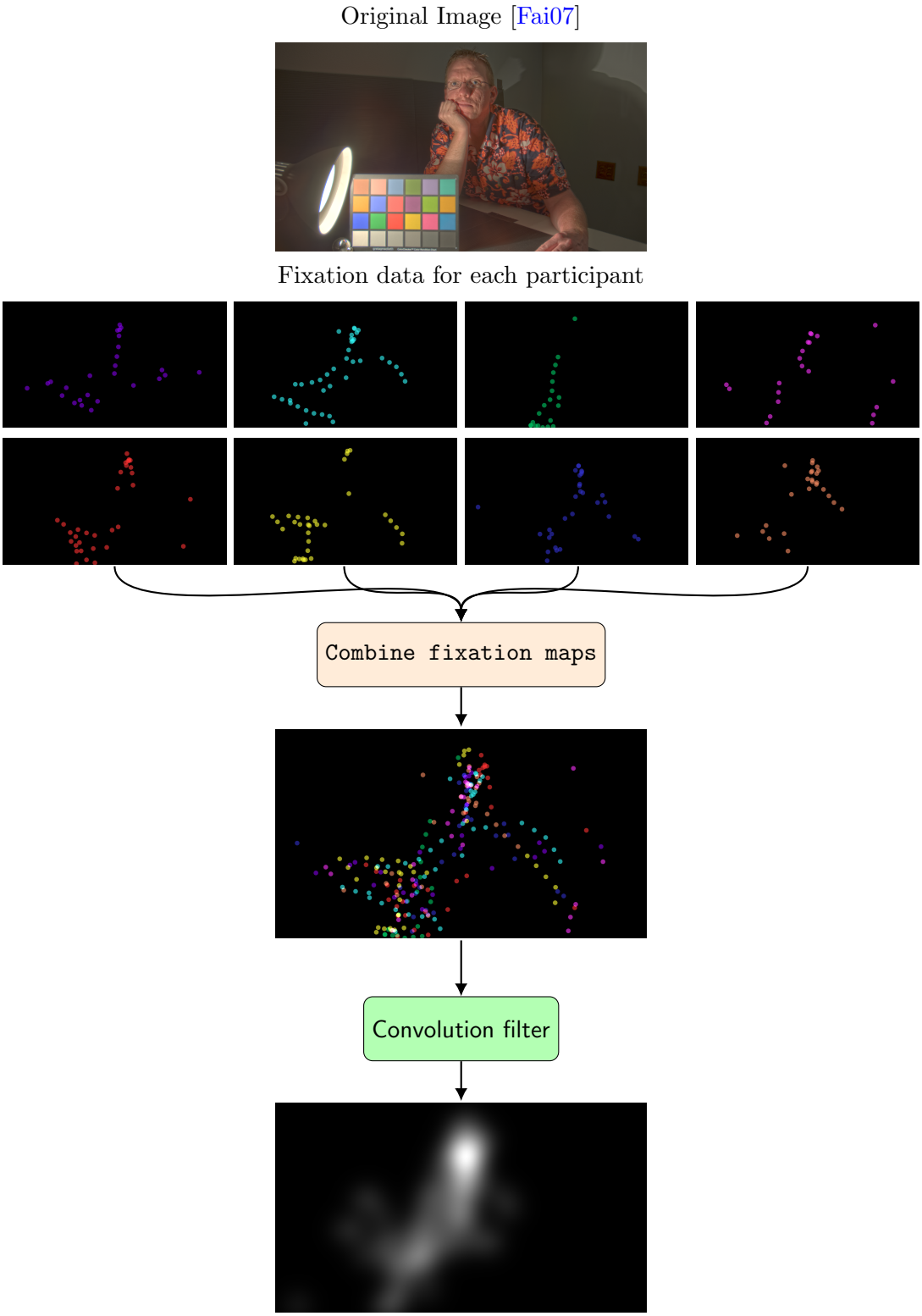


Figure 6.5: An illustration of the fixation map creation process for the HDRMark [Fai07] image

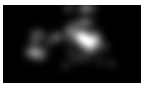
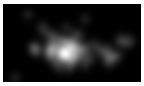
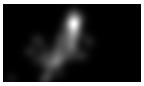
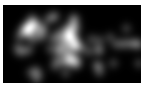
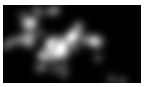
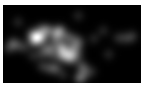
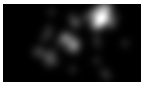
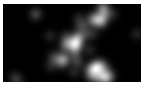
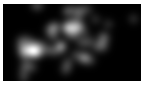
	Tonemapped image	500	1000	2000	4000
507					
BarHarborSunrise					
CemeteryTree(1)					
GoldenGate(1)					
HDRMark					
OldFaithfulInn					
RITTiger					
SmokyTunnel					
TheNarrows(2)					
TunnelView(2)					

Table 6.1: Fixation Density Maps for ten scenes at various brightnesses

in these graphs, there is a general downwards trend as brightness increases.

6.3.3 Self Prediction

Section 6.3.2 provided insight into the similarity of LDR to fixation maps of increasing brightness, whereas this section aims to analyse the reliability of fixation maps at varying luminance levels. Reliability, in this context, refers to the uniformity of fixations from various observers and therefore their likelihood of predicting new fixations, under similar conditions, accurately.

Comparing a condition against itself requires splitting the fixations for each image and condition into multiple subsets, as the metrics used in this study operate in a pairwise manner. k-fold cross validation, is one method for achieving this comparison. Referring to Figure 6.6, which illustrates the division of subsets, this functions by initially selecting a subset size, k, where k is divisible by the total number of participants, n_p. Participants' fixation data are then randomly assigned into n_p/k equal groups. Following this the smaller group, of size k, is kept as a validation group while the remaining $n_p - k$ participants are put aside as a training group, which is used to predict the validation group's fixations. This process is repeated for all n_p/k groups, then the results of each test is averaged to find the true result. Thus every group is compared against itself, this is represented visually in Table 6.5.

Table 6.6 shows the results of a repeated measures ANOVA across brightnesses for each metric. For this test a subgroup of four viewers was chosen as the validation set, and eight for the 'training' set. The main effect was not found to violate the the assumption of sphericity ($p > 0.05$), with the exception of the EMD metric, for which

Screen Brightness (cd/m^2)			
500	1000	2000	4000
A	B	C	D

Table 6.2: Participant groups for LDR baseline analysis

	Initial fixation (2s)		Free viewing (10s)	
Metric	$F(2, 158)$	p value	$F(2, 158)$	p value
AUC	10.33	< 0.01	7.28	< 0.01
EMD	2.86	> 0.05	2.30	> 0.05
SIM	6.44	< 0.01	1.84	> 0.05
CC	10.91*	< 0.01*	4.39	< 0.05
NSS	13.57	< 0.01	6.07	< 0.01
KLD	2.42*	> 0.05*	3.53	< 0.05

Table 6.3: Results of inter-brightness ANOVA

	Initial fixation (2s)	Free viewing (10s)
Metric	Brightness	Brightness
AUC	1000 2000 4000	2000 1000 4000
EMD	2000 1000 4000	2000 1000 4000
SIM	1000 2000 4000	2000 1000 4000
CC	1000 2000 4000	2000 1000 4000
NSS	1000 2000 4000	2000 1000 4000
KLD	1000 2000 4000	2000 1000 4000

Table 6.4: Differences in inter-brightness comparisons, significant at $p < 0.05$

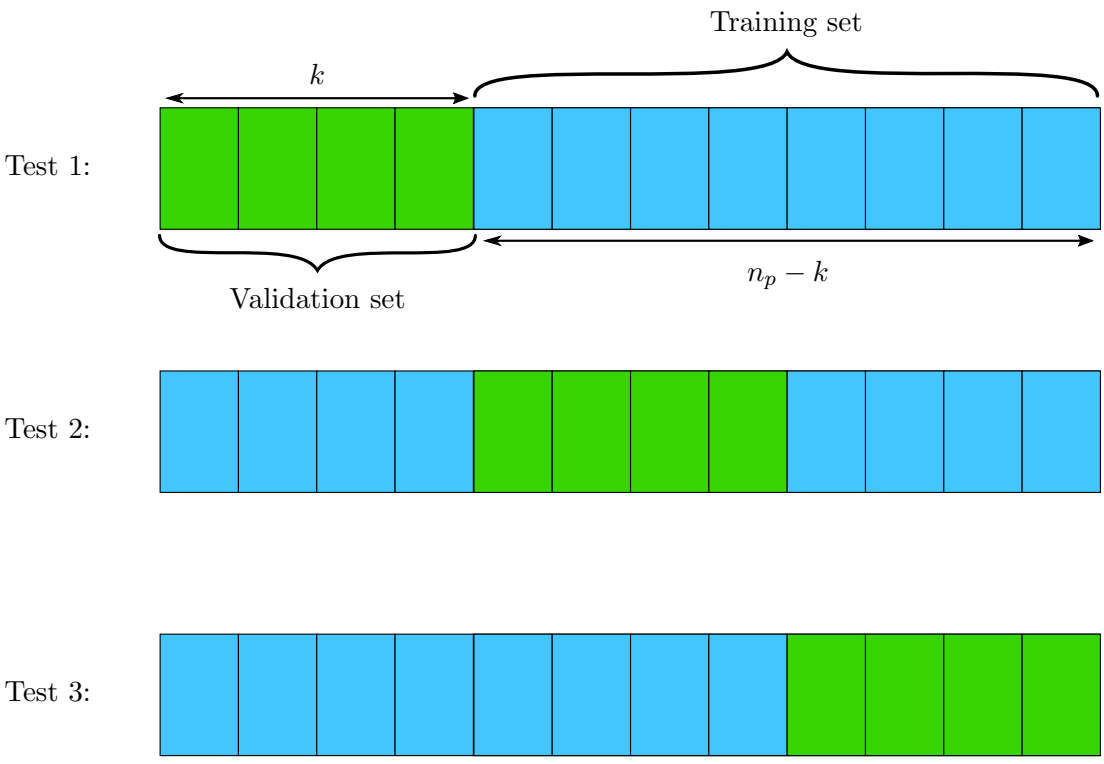


Figure 6.6: A visualisation of the k-fold algorithm used in this study

the Greenhouse-Geisser corrected values are reported and marked in the table with an asterisk. The degrees of freedom for the F-score in this case being $F(2.71, 214.01)$.

The results suggest a decrease in the ability of fixation maps to predict new maps, as brightness increases. This result was found to be significant in four out of the six metrics used. The results of the Bonferroni-corrected pairwise comparisons can be seen in Table 6.7 and Figure 6.7, 6.8, 6.9, 6.10, 6.11 and 6.12, the self prediction scores are labelled as "Self Prediction 2s" for the initial fixation and "Self Prediction 10s" for the full free viewing period. Once again we see a downwards trend in performance as brightness increases.

6.3.4 LDR as a baseline predictor

The previous analyses are focused mainly on addressing Hypothesis 1, as stated in section 6.2.1, examining the reliability of fixations with varying screen luminances levels. In order to address Hypothesis 2 an additional layer of analysis is needed. The metrics used in this chapter are pairwise, the statistical analyses that are conducted herein are always contrasts of similarities between groups of fixations. Thus to establish whether screen brightness has affects the viability of utilising a set of fixations as a baseline, the contrast conducted is between each conditions ability to predict itself, as in Section 6.3.3 and the ability of LDR to predict the same condition. This comparison is represented visually in Table 6.8.

Using the same validation and training set sizes as in Section 6.3.3, yielded the results shown in Table 6.9 and Figure 6.7, 6.8, 6.9, 6.10, 6.11 and 6.12. In these graphs the initial fixation and free viewing periods are labelled as "Self vs. LDR 2s" and "Self

Screen Brightness (cd/m^2)			
500	1000	2000	4000
A	B	C	D

Table 6.5: Participant groups for self prediction analysis

	Initial fixation (2s)		Free viewing (10s)	
Metric	$F(3, 237)$	p value	$F(3, 237)$	p value
AUC	7.37	< 0.01	7.74	< 0.01
EMD	3.29^*	$< 0.05^*$	2.37	> 0.05
SIM	8.81	< 0.01	9.94	< 0.01
CC	9.78	< 0.01	11.44	< 0.01
NSS	10.31	< 0.01	8.59	< 0.01
KLD	6.34^*	$< 0.01^*$	7.97	< 0.01

Table 6.6: Self prediction results of repeated measures ANOVA

	Initial fixation (2s)	Free viewing (10s)
Metric	Brightness	Brightness
AUC	500 1000 2000 4000	500 2000 1000 4000
EMD	1000 2000 500 4000	500 1000 2000 4000
SIM	500 1000 2000 4000	500 1000 2000 4000
CC	1000 500 2000 4000	500 1000 2000 4000
NSS	1000 500 2000 4000	500 1000 2000 4000
KLD	500 1000 2000 4000	500 1000 2000 4000

Table 6.7: Self-prediction pairwise significant differences at $p < 0.05$

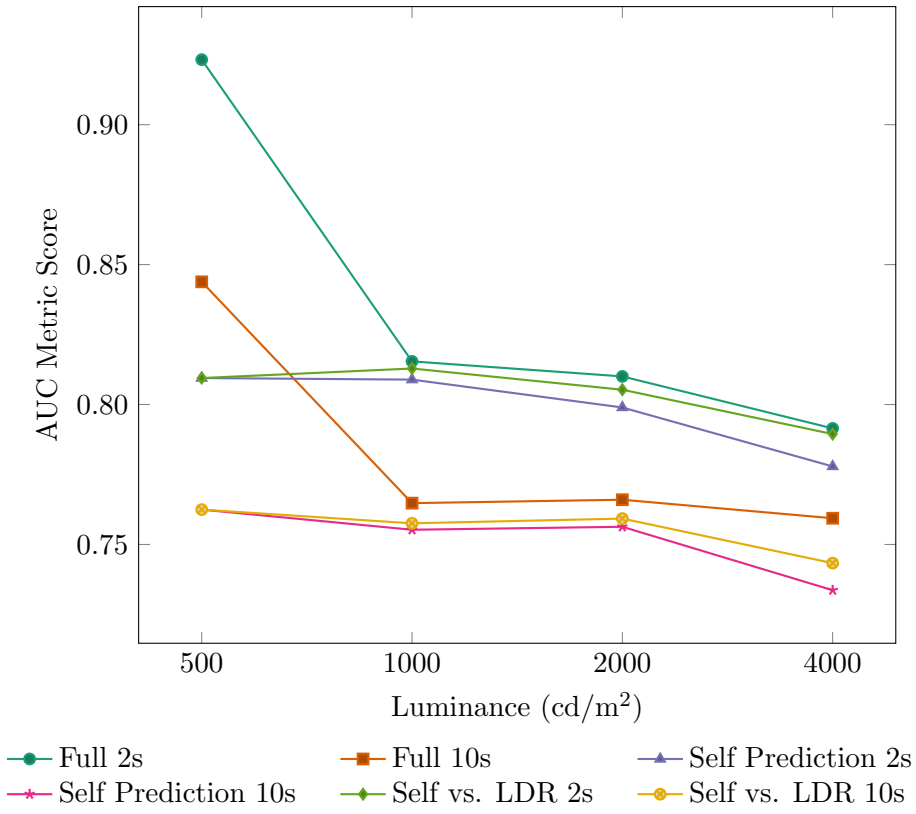


Figure 6.7: The results of fixation map comparisons for the AUC metric

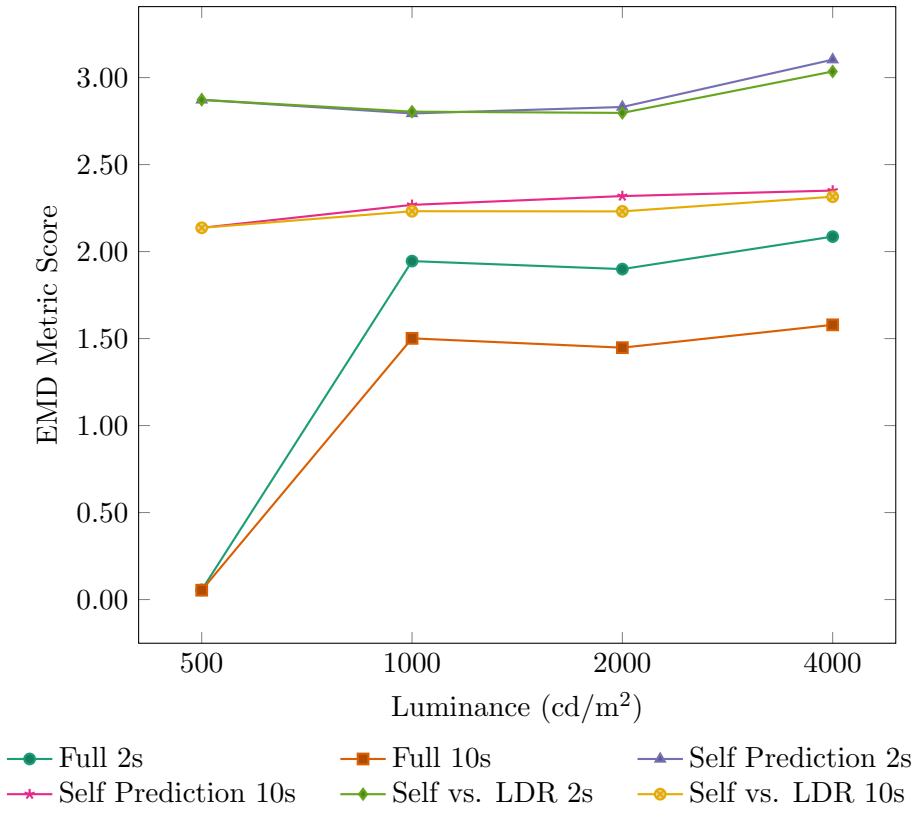


Figure 6.8: The results of fixation map comparisons for the EMD metric

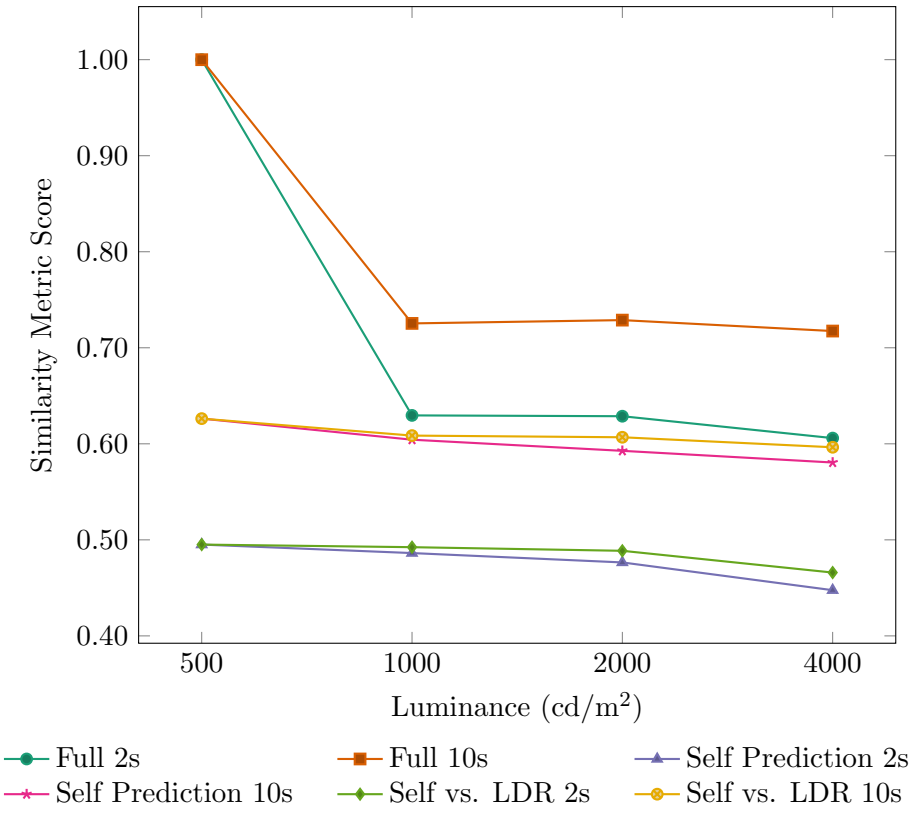


Figure 6.9: The results of fixation map comparisons for the Similarity metric

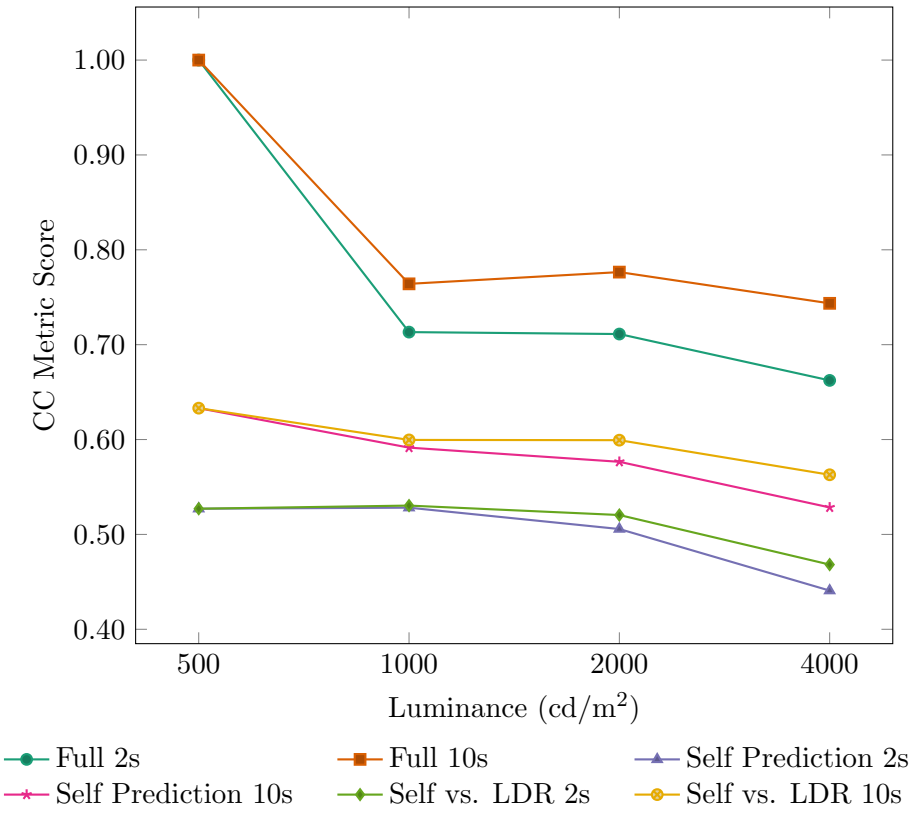


Figure 6.10: The results of fixation map comparisons for the CC metric

vs. LDR 10s", respectively. For five out of the six metrics, for both the initial fixation and full viewing period, the results suggest a significant difference in the predictive power.

Screen Brightness (cd/m^2)			
500	1000	2000	4000
A	B	C	D

Table 6.8: Participant groups for LDR baseline analysis

	Initial fixation (2s)		Free viewing (10s)	
Metric	$F(1, 79)$	p value	$F(1, 79)$	p value
AUC	7.21	< 0.01	7.70	< 0.01
EMD	0.56	> 0.05	3.16	> 0.05
SIM	9.48	< 0.01	14.60	< 0.01
CC	4.37	< 0.05	13.94	< 0.01
NSS	5.71	< 0.05	14.80	< 0.01
KLD	7.84	< 0.01	16.63	< 0.01

Table 6.9: The results of a repeated measures ANOVA analysis of self and inter-brightness prediction

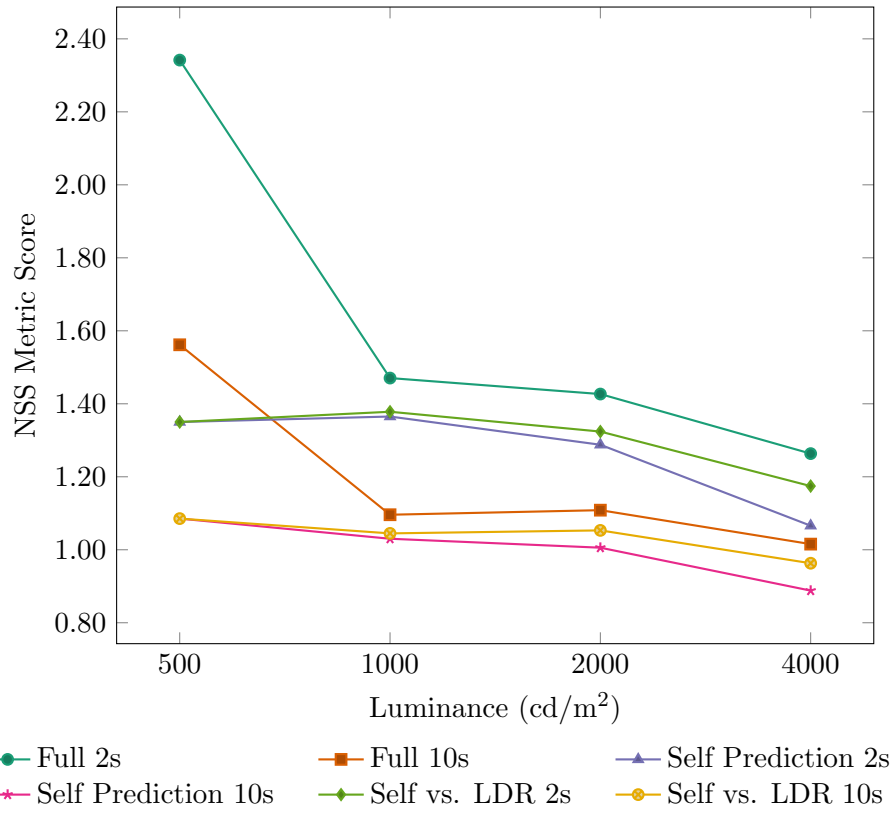


Figure 6.11: The results of fixation map comparisons for the NSS metric

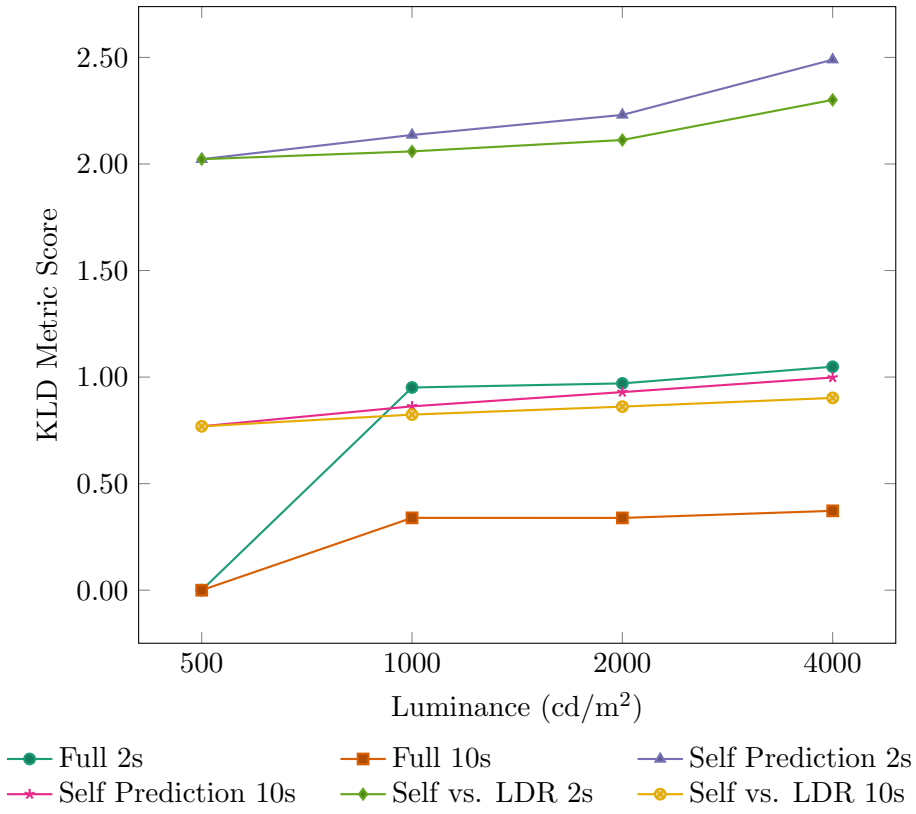


Figure 6.12: The results of fixation map comparisons for the KLD metric

6.4 Analysis

Results of the inter-brightness comparison, as seen in Table 6.3, suggest that during the initial fixation period the variation in fixation patterns across brightnesses was significant. This result is based on the findings of four out of the six metrics, with EMD being the exception. The effect decreased in size in the full ten second free-viewing period, with only four out of the six metrics finding a significant effect. Similarity and EMD found no significant effect.

Pairwise comparisons for the initial fixation found significant differences, represented visually in Table 6.4, between the results of the six metrics, excluding EMD and KLD, for a peak brightness of 4,000 cd/m^2 and the results for 1,000 and 2,000 cd/m^2. This lends credence to the idea that initial fixations might be affected by increasing brightness. The pairwise comparisons, also in shown in Table 6.4, of metrics for the full ten seconds free-viewing period found fewer significant differences between brightness levels, with the EMD, Similarity, and KLD metrics finding no difference between any levels. In all tests the brightest group showed the greatest divergence from the original LDR images, which also suggests an effect of brightness on the validity of LDR datasets for predicting HDR fixations and is supportive of Hypothesis 1.

Table 6.6 shows the results of the self-prediction tests. For the initial fixation, all six metrics found the effect of brightness to be significant. For the longer free viewing period, significant differences were found in all but the EMD metric results. The results of all significant scores suggested a decrease in the predictive power of fixation maps as the brightness of the images increased. This result is interesting as it implies that as screen or image brightness increases there is a significant change in the uniformity of observers viewing strategies, lending further credence to Hypothesis 1. Though it is likely that, given an increase in the available visual information, images with higher dynamic ranges require more observers to obtain reliable eye-tracking data. Further study is needed to confirm or refute this.

The results of pairwise comparisons of the self prediction tests can be seen in Table 6.7. The results for the initial fixations show clearer groupings than their counterparts in the free viewing period, with all six metrics finding some groupings in the initial fixation

period. Pairwise tests of conditions within the free viewing period found groupings in five of the six metrics, with EMD being the exception. This lends further credence to the idea that eye-tracking data is less useful for predicting new fixations as brightness increases.

The comparison of the within- and inter-brightness evaluations in Table 6.9 implies that, for five out of the six metrics, fixations from LDR images were significantly better predictors of new HDR fixations than those from HDR images. This is strong evidence that the assertion made in Hypothesis 2 is false. This could also be caused by a wider variation in viewing strategies for HDR images, however further study is recommended.

6.5 Conclusion

This chapter details the procedure and results of an eye-tracking study of images at varying peak brightnesses, with a view to determining if models of visual attention and selective rendering methods should be adapted to target the peak luminance of a given display. The obtained fixation data was analysed using six common metrics used to rank saliency models in the MIT saliency benchmark [BJB+]. The analysis provided evidence that the predictive power of LDR fixation data, as well as the uniformity of viewing strategies across observers, decreases significantly as the brightness of the displayed image increases, suggesting a variation in viewing strategies between HDR and LDR data. However it was also observed, through a within-brightness k-folds cross validation, that as the brightness of a stimulus increases, the predictive power of eye-tracking data collected from it decreases significantly. In addition the data suggests that this power is lower than that of the inter-brightness predictive power. This could be due to a larger variation in viewing strategies for HDR images, but data from a larger number of participants would need to be collected to fairly examine this hypothesis.

The results imply that HDR eye-tracking datasets are not better predictors for new fixations on HDR stimuli, which draws into question the necessity of HDR eye-tracking datasets, though further study is needed.

Chapter 7

Ambient Light, Visual Attention and Visual Fatigue

Subjective evaluations of imaging techniques, such as tone mapping operators, video compression algorithms, down sizing algorithms, and eye tracking data collection often occur in laboratory conditions, with very low ambient light levels. With the recent advent of HDR screens, such as the SIM2 HDR display [SIM15], capable of outputting luminance reaching well into the photopic range, it has become necessary to assess accommodation effects on visual attention. Currently no research has been done on the effects of ambient illuminance levels on visual attention, neither in standard models nor in the small number of HDR saliency models which have emerged in recent years, and which have been discussed in previous chapters. This is pressing as the human visual system is only capable of perceiving detail concurrently in luminance ranges with ratios of $1 : 10^4 \text{cd/m}^2$ [BADC17], without adaptation.

The primary focus of this chapter is to establish whether ambient illuminance is a significant factor on visual attention for standard and high dynamic range content, while providing a eye-tracking dataset which can be used to validate futures models. This builds on the work of the previous chapters, by expanding the empirical basis for models of visual attention for HDR content and enabling further targeting of selective rendering algorithms.

Furthermore, extensive research has been conducted on the effect of e-readers

[BCDZ+14], normal televisions [IKK+09], 3D-TVs [LHP10], and computer monitors [LPWM09] on visual fatigue. Previous work has shown that display luminance and ambient light have an effect on the discomfort of viewers [BCDZ+14, SSH05]. Despite the recent emergence of a number of high luminance display technologies there have been no studies on the effect of HDR content on visual fatigue under varying ambient light conditions, though a recent study by Yang et al. [YJL17] determined that for visual fatigue increased with screen luminance for mid-range HDR displays ($<$1,500 cd/m^2. This chapter seeks to rectify this by analysing the blink rates and subjective self-assessments of observers' after viewing LDR and HDR content under varying environmental conditions.

7.1 Outline

This chapter seeks to assess what relation, if any, image brightness and ambient illumination have with visual attention and visual fatigue. These factors are assessed independently through an eye tracking experiment.

The structure of this chapter is as follows, first in the next section the method of the eye tracking experiment, including its design, materials and procedure are presented. The subsequent section outlines the format of the resulting data and statistical tests. Finally a discussion of these results and their context is presented.

7.2 Methodology

This section outlines both an objective eye tracking data collection for LDR and HDR image content under varying lighting conditions, and a psychophysical test, wherein participants were asked to rate their visual fatigue on a seven point Likert scale after the image sequence completed. The purpose and design of the experiment is detailed first, then the procedure for generating the materials is outlined. Finally an overview of the experimental procedure and demographic make-up of the participants is described.

Figure 7.1: A sample of images used in this experiment

7.2.1 Design

The experiment outlined in this section aims to provide sufficient data to test the following two hypotheses:

Hypothesis 1 (H1): *The brightness of a display and the ambient illuminance of the environment will have a significant effect on visual attention.*

Hypothesis 2 (H2): *The brightness of a display and the ambient illuminance of the environment will have a significant effect on visual fatigue.*

To this end, eye tracking data is collected for sixty LDR and HDR images with a resolution of 1920×1080, taken from the Fairchild database [Fai07], Technicolor's HDR dataset [LLF13], and the dataset of Froehlich et al. [FGE+14]. This provided sufficient material for an analysis of the effects of ambient light and screen brightness on visual attention for two view durations, an initial fixation period of 2 s and a free-viewing period of 12 s. The change to 12 s from the previous chapter is in part to bring the test in line with previous HDR eye tracking studies [NKHE15], and in part to achieve a 15 minute total viewing time. Previous works [NKHE15, DNPN14, BPT10] have collected

	Ambient Light Level (lx)		
Brightness	0	250	1500
500	A (8)	B (8)	C (8)
8000	D (8)	E (8)	F (7)

Table 7.1: Participant groups for each condition, the number of participants in each group is shown in brackets

data in dark laboratory environments, though the effects of dark environments the validity of HDR eye tracking datasets is so far undocumented.

The independent variables that will be manipulated during the eye tracking experiment are peak screen *brightness*, which has two levels (500 and 8,000 $\mathrm{cd/m^2}$) and the *illuminance* incident on the screen, which has three levels (0, 250 and 1,500 lx), the experiment therefore has a 2×3 factorial between participants design. The mid level ambient illuminance is per the recommendations in ITU-R BT.500-13 [IR12] and BT.2022 [Rec12], which recommends that the environmental illumination be equal to 15 of the peak screen luminance or 200 lx for home environments when using screens with peak luminances of 70 to 500 $\mathrm{cd/m^2}$. The specification could not be met entirely and 250 lx was the closest approximation achievable in the experimental environment. The high illuminance level is taken from Melo et al. [MBB+15], and represents the light level recorded outside on a sunny day, it is also roughly in line with the guidelines mentioned previously in ITU-R recommendations BT.500-13 and BT.2022 for 8,000 $\mathrm{cd/m^2}$, which would be 1,200 lx. These recommendations were also used in the HDR MPEG standardisation [LFH15]. Both variables were manipulated in a between-participants design, with a total of six participant groups, one for each of the six conditions. Forty-eight are then divided into groups evenly across the conditions as in Table 7.1, each participant sees every image, the order of which is randomised to eliminate potential confusing effects.

The dependent variables, which are being examined are fixation patterns and blink rates of observers, which will be captured using an eye tracking software package and a

subjective self-assessment of eye fatigue, which will be evaluated using a survey following the study. The fixation data variable is analysed across images, in a repeated measures design, as the independent variables are manipulated while the displayed images remain constant. The total sample size for this variable is sixty images, a sample of which can be seen in Figure 7.1. In contrast visual fatigue is analysed between participants, with forty-eight participants in total.

The distance to the screen was kept constant at 1.8 m in accordance with ITU recommendation 2022 [Rec12]. The viewing time for each image was also held constant, at 12 s in line with previous research [NKHE15, JDT12]. Between each image a grey spacer image was displayed for 3 s with a black dot in the centre, on which participants were asked to focus, to avoid their eyes accommodating to the environment between images. This has the drawback that the first fixations of participants are more likely to be in the centre of the screen, however in the previous experiment, where no dot was used some participants did not look at the screen during the initial 2 s viewing period. Therefore on a balance of factors it was decided to include the grey dot on the spacer image. The eye tracking device was 50 cm from each participant, and was calibrated with a five point calibration procedure before recording began.

7.2.2 Materials

The experiment took place in a room illuminated by a calibrated light source, the images were viewed on a 47 inch SIM2 HDR monitor [SIM15]. In addition to this the eye-tracker was raised 10 cm above the table and secured with clamps. The stereo-head containing the cameras for eye-tracking was attached to a laptop, which was placed on a nearby table. See Figure 7.3 for an illustration of the apparatus.

The HDR images used in the experiment had a resolution of 1920×1080 and taken from the Fairchild database [Fai07], Technicolor's HDR dataset [LLF13], and the dataset of Froehlich et al. [FGE+14]. Prior to display they were linearly scaled such that the 99^{th} percentile of image brightness was at the desired brightness level, while Krawczyk et al. [KMZS07] observed in their paper that this scaling method does not produce high levels of subjective image quality compared to some other methods, it ensures that

Figure 7.2: The Experiment Environment

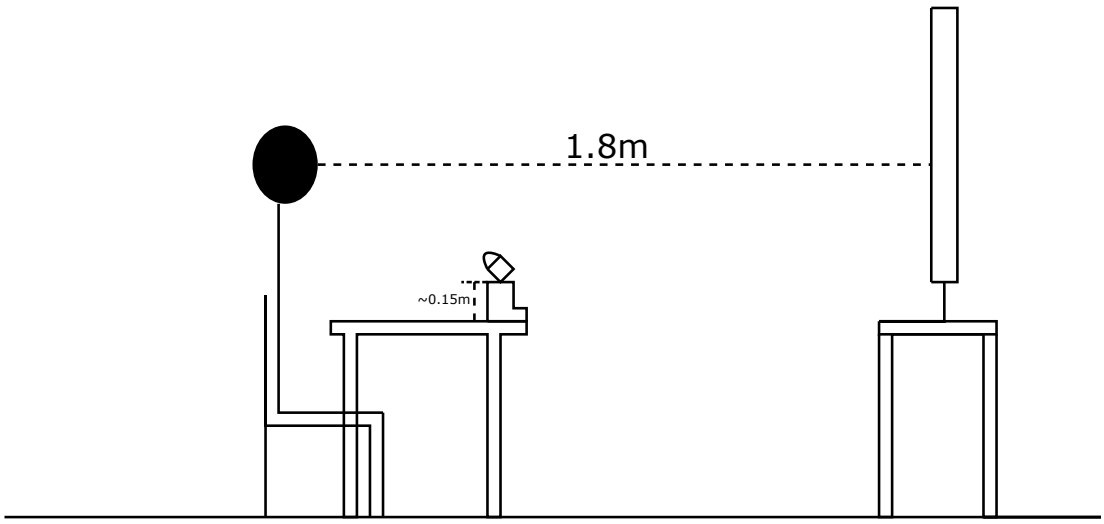


Figure 7.3: Experimental Setup

brightness levels across images are consistent and is not affected by image content.

The eye tracker used in the study was the SMI-Red [Ins17] camera and iViewX software package, from SensoMotoric Instruments, it has a resolution of $-/+0.5°$, a framerate of 60 Hz was used for testing. In order to facilitate the efficient subdivision and uniform treatment of captured data, raw data of event statistics were exported from the eye tracking software package for each participant and processed independently rather than utilising the functionality of the software. This was necessary for a number of reasons, primarily for the sorting and control of data. The iViewX software suite is not compatible with HDR images, which necessitated the use of custom software, running on the same computer. Therefore the system had no knowledge of which image was shown, at what point and for how long. The start and end time for the display of each image was recorded in the image display software, and that time was matched against the fixation events to create fixation maps. The full MATLAB code that was used to process and analyse these results and those of the previous chapter are available here: `https://bitbucket.org/badger-tennis/thesis_utility/src/master/`

7.2.3 Procedure

Upon entering the test environment participants were read a prepared script, informing them of the nature and purpose of the experiment, to avoid task dependent effects the script was the same for every participant, they were then asked for their informed consent. Following this, participants were asked to complete a five point calibration and validation sequence. Participants then viewed a fifteen minute the image sequence under a single ambient illuminance and image brightness condition. In order to encourage participants to remain attentive for the full duration of the sequence, they were informed that there would be a short memory test following the image sequence.

After the completion of the image sequence subjects were asked to complete a qualitative self assessment of visual fatigue. This took the form of a survey based on the work of Heueur et al. [HHKR88], wherein six statements are rated for agreement on a Likert scale from 1 to 7. In order to avoid priming the participants with a task, the questionnaire was only given to the participants after the experiment, as

opposed to taking measurements before and after the experiment. The statements which participants were asked to rate their agreement with are as follows:

1. I have difficulties in seeing.
2. I have a strange feeling around the eyes.
3. My eyes feel tired.
4. I feel numb.
5. I have a headache.
6. I feel dizzy looking at the screen.

A reading from a humidity meter was taken prior to each trial for posterity, as humidity has an effect on eye-strain [BVK+05]; the average of these readings was 21%, there was no variance due to an air conditioning.

7.2.4 Participants

Forty-eight university students, staff, and employees took part in the experiment. One participant had to be discarded during the analysis phase, owing to an error with the eye-tracker which prevented data collection in the latter half of the image sequence. Of the remaining forty-seven participants, seven female, and forty male. All participants reported having normal or corrected to normal vision. The average reported age of all the participants was 26.7 years. Nine of the participants had some knowledge of visual attention or HDR imaging.

7.3 Processing Fixation Data

For every participant, fixation locations were recorded in image space, or pixel coordinates. The SMI-Red eye tracker [Ins17] used in this study, flags fixations, blinks, and saccades automatically. Locations fixated during saccades contain no relevant attention data, as the brain doesn't process imagery during a saccade, [EMZ09, JDT12] and are

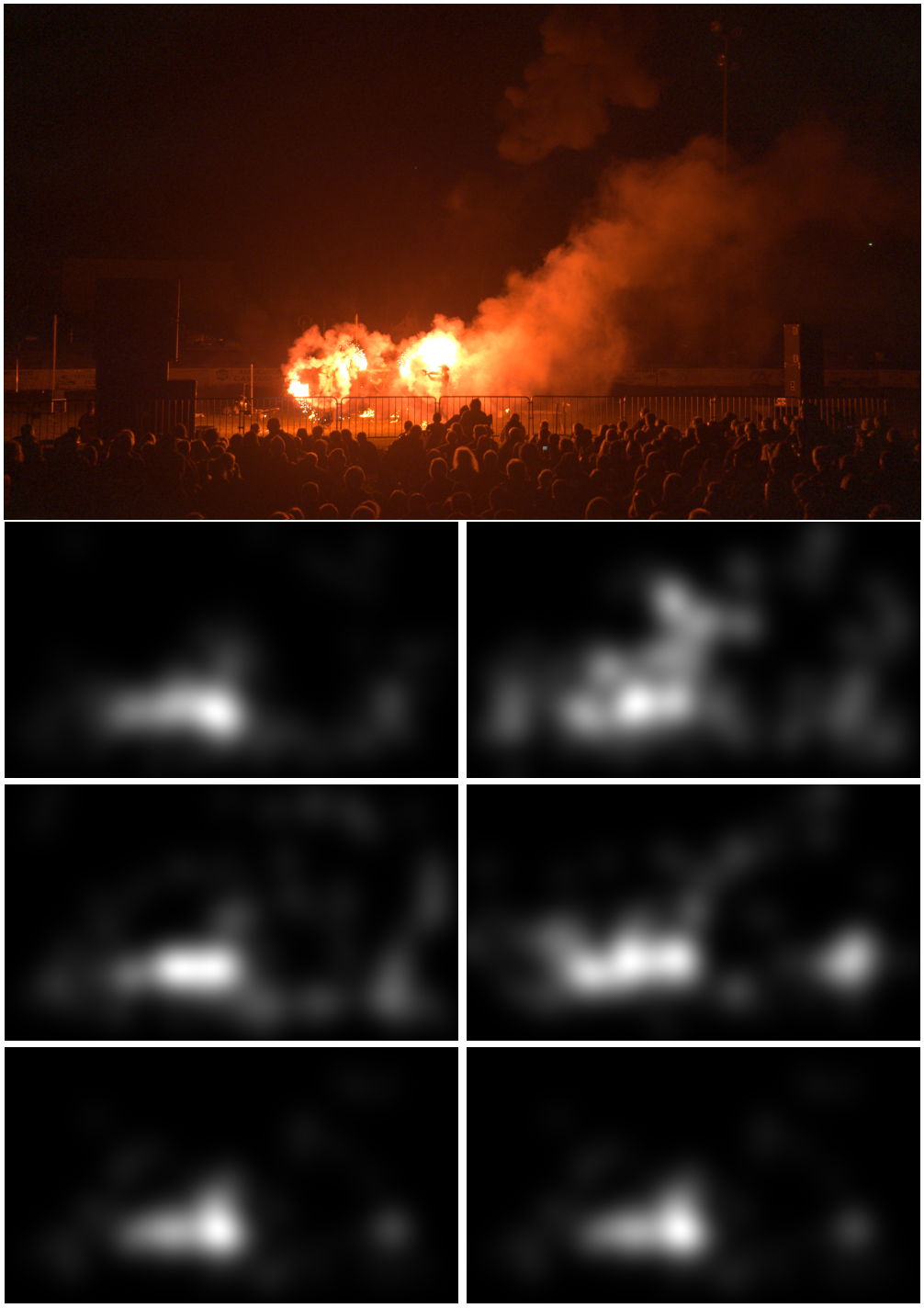


Figure 7.4: An example of fixation density maps for an image used in the experiment. The left column contains the LDR fixation maps, with ambient light increasing from top to bottom. The right column contains respective HDR maps

therefore discarded, along with blinks. The fixations are then grouped by image for each participant, by matching the start and end times for which an image was shown, taken from the image display software, with the local time recorded by the eye tracker. Fixation locations are then stored per image and per condition, as per the groupings in Table 7.1. Fixation maps and fixation density maps are then created as per Algorithm 1, as in Chapter 6. A sample of the fixation maps for each condition for a single image can be seen in Figure 7.4.

To quantify the differences between these fixations, the six metrics employed in Chapter 6 and discussed in detail in Chapter 3 are utilised, specifically; area under the curve (AUC), the Earth Mover's Distance (EMD), the Similarity metric, Pearson's Correlation Coefficient (CC), Normalised Scanpath Saliency (NSS), and symmetric Kullback Leibler Divergence (KLD).

Visual fatigue is examined through two factors, blink rates, higher rates of which indicate higher levels of eye-strain [BCDZ+14] and the survey responses.

7.4 Results

This section presents a summary of the key results from the experiment described in the previous section, and where relevant a description of how they were processed. The structure of the following results sections is as follows; Section 7.4.1 outlines an analysis of the within-brightness variation of viewing strategies with respect to ambient illumination; Section 7.4.2 analyses how the similarity of viewing strategies on LDR and HDR content vary as brightness increases; Section 7.4.3 analyses the overall effect of both brightness and ambient illuminance on the uniformity of fixations within participant groups; Section 7.4.4 evaluates the effectiveness of using LDR eye tracking data captured in dark environments to evaluate viewing strategies across the brightness and ambient light condition; finally, Section 7.4.5 offers an analysis of how the self-reported visual fatigue and blinking rates vary between participants across all conditions.

	Ambient Light Level (lx)		
Brightness	0	250	1500
500	A	B	C
8000	D	E	F

Table 7.2: Participant groups for analysis of the effects of ambient light within brightness levels

7.4.1 Ambient Illumination and Visual Attention

In order to examine the effect of ambient light on fixation patterns, metrics are applied across the ambient variable for each illumination. Table 7.2 represents this visually. By contrasting each between each condition of the ambient illuminance variable, within brightnesses a sense is formed of the relation between ambient light and fixation patterns. Table 7.3 and 7.5 show the results of 1×3 factorial ANOVA of these comparisons, for LDR and HDR content, respectively. Where the assumption of sphericity was violated F-scores are marked with an asterisk and the more conservative, Greenhouse-Geisser corrected F-scores and p-values are reported. Pairwise comparisons contrasting the similarities across ambient illuminance levels are shown in Table 7.4 and 7.6 for LDR and HDR brightnesses, respectively. Wherein, contrasts are arranged in order of decreasing mean similarity from left to right.

The results for five of the six metrics show significant results for both viewing periods for the LDR ANOVA, with mean similarities between conditions decreasing as ambient illuminance increases. The HDR results show a similar level of significance, though sphericity was violated for multiple metrics. In addition, the opposite trend in mean similarities was observed, with scores increasing with increasing ambient illuminance. This is supportive of Hypothesis 1.

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	2.88	0.060	22.9	0.001
EMD	6.09	0.003	3.15	0.046
SIM	8.79	0.001	12.3	0.001
CC	6.13	0.003	14.6	0.001
NSS	7.47	0.001	17.3	0.001
KLD	7.51	0.001	8.35*	0.001

Table 7.3: The results of a one way ANOVA across ambient illuminance levels for LDR images

	Initial fixation (2s)			Free viewing (10s)		
Metric	Pairing			Pairing		
AUC	$A \times B$	$B \times C$	$A \times C$	$A \times B$	$A \times C$	$B \times C$
EMD	$A \times B$	$A \times C$	$B \times C$	$A \times B$	$A \times C$	$B \times C$
SIM	$A \times B$	$A \times C$	$B \times C$	$A \times B$	$A \times C$	$B \times C$
CC	$A \times B$	$A \times C$	$B \times C$	$A \times B$	$A \times C$	$B \times C$
NSS	$B \times C$	$A \times B$	$A \times C$	$A \times B$	$B \times C$	$A \times C$
KLD	$A \times B$	$A \times C$	$B \times C$	$A \times C$	$A \times B$	$B \times C$

Table 7.4: Pairwise comparisons for LDR images across ambient illuminance levels, significant at $p < 0.05$

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	5.86	0.007	8.58	0.001
EMD	2.06*	0.133	7.00*	0.002
SIM	16.6*	0.001	20.8*	0.001
CC	19.3*	0.001	15.3*	0.001
NSS	28.9	0.001	32.4*	0.001
KLD	3.63*	0.044	15.82*	0.001

Table 7.5: The results of a one way ANOVA across ambient illuminance levels for HDR images

	Initial fixation (2s)			Free viewing (10s)		
Metric	Ambient Illuminance			Ambient Illuminance		
AUC	$E \times F$	$D \times F$	$D \times E$	$D \times F$	$E \times F$	$D \times E$
EMD	$D \times F$	$E \times F$	$D \times E$	$E \times F$	$D \times F$	$D \times E$
SIM	$E \times F$	$D \times F$	$D \times E$	$E \times F$	$D \times F$	$D \times E$
CC	$E \times F$	$D \times F$	$D \times E$	$E \times F$	$D \times F$	$D \times E$
NSS	$E \times F$	$D \times F$	$D \times E$	$D \times F$	$E \times F$	$D \times E$
KLD	$E \times F$	$D \times F$	$D \times E$	$E \times F$	$D \times F$	$D \times E$

Table 7.6: Pairwise comparisons for HDR images across ambient illuminance levels, significant at $p < 0.05$

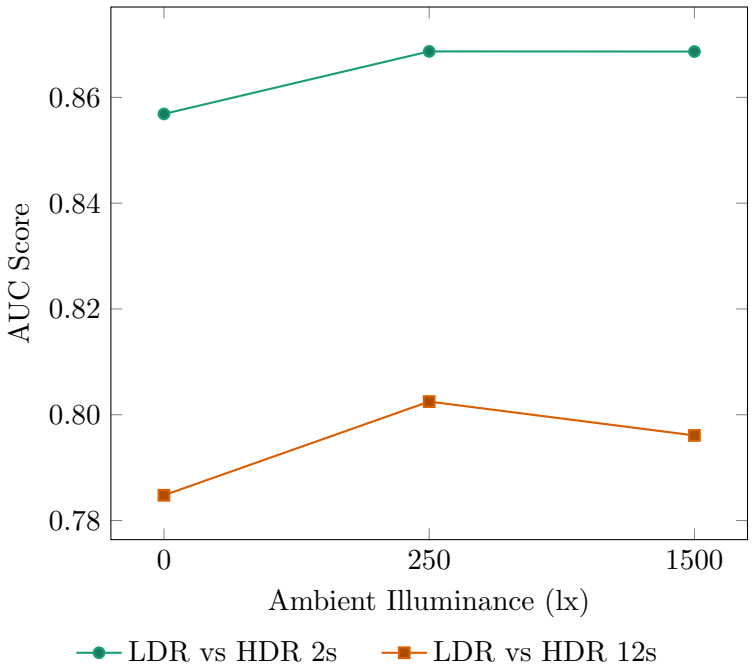


Figure 7.5: Mean AUC scores comparing LDR vs HDR fixations for varying ambient illumination levels

7.4.2 Ambient Illumination and LDR vs. HDR

This section provides an analysis on relationship between LDR and HDR fixations under varying ambient illuminance levels.

Results from the previous chapter suggest that the similarity of fixations decreases as the luminance of an observed image increases. In order to evaluate the effect of ambient brightness on this relationship, metrics are applied across brightness levels, where LDR fixations used to predict HDR fixations. Results of the metrics are then compared in a one way ANOVA across ambient illuminance levels. Table 7.7 illustrates these comparisons visually. Each group contains the fixation data of eight participants.

The results of the ANOVA are displayed in Table 7.8. Values where $p < 0.05$ represent analyses where a significant effect of ambient light on the similarity of LDR and HDR fixations was observed. As is visible in the table, the effects were found to be more significant for the longer free viewing fixation time of 12 s, with all metrics observing a significant effect, save for the symmetric Kullback Leibler Divergence (KLD)

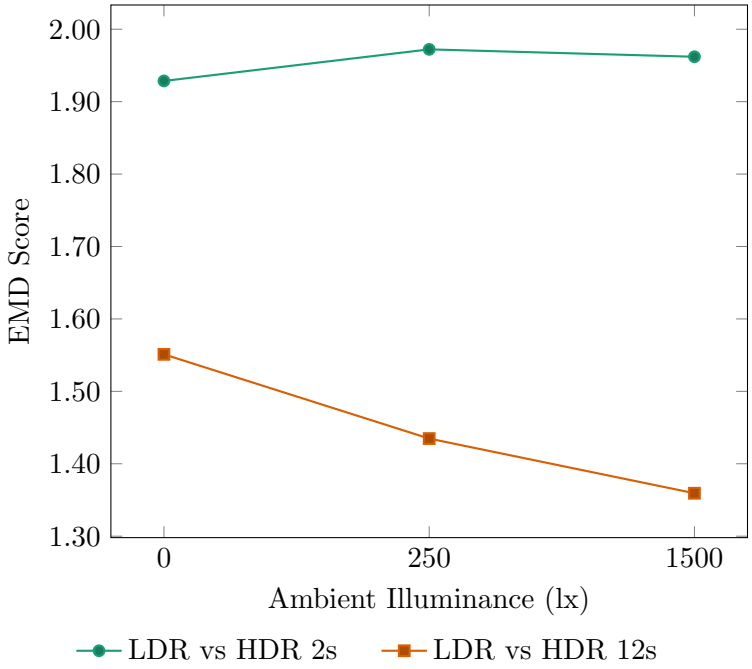


Figure 7.6: Mean EMD scores comparing LDR vs HDR fixations for varying ambient illumination levels

	Ambient Light Level (lx)		
Brightness	0	250	1500
500	A	B	C
8000	D	E	F

Table 7.7: Participant groups for an analysis of the effects of ambient light on the similarity of LDR and HDR fixation maps

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	2.66	0.074	7.11	0.001
EMD	0.07	0.932	4.10	0.019
SIM	3.31	0.040	5.62	0.005
CC	4.29	0.016	5.63	0.005
NSS	18.7	0.001	13.3	0.001
KLD	0.04	0.964	2.96	0.056

Table 7.8: Results of an ANOVA comparing LDR fixations similarity to HDR fixations with increasing ambient illuminance

metric.

The means of the resulting metric scores for each ambient illuminance can be seen in Figure 7.5 7.6 7.7 7.8 7.9, and 7.10. Note that a higher score indicates a greater similarity for the LDR and HDR fixation groups, with the exception of the Earth Mover's Distance (EMD) and KLD scores wherein a lower score indicates a greater similarity.

7.4.3 Ambient Illumination, HDR Content and Visual Attention

The intrinsic value of an eye tracking dataset lies in its consistency, the more representative it is of the population of potential viewers, the more useful it is for model validation and training. Assessing the variation in this consistency across brightness and illuminance requires that contrasts be conducted within conditions. This is accomplished through k-fold cross validation, wherein every group is compared against itself, this is represented visually in Table 7.10, and was previously explained in Section 6.3.3.

For the following analysis a value of $k = 4$ was chosen, which leaves only two groups of four and two iterations for each condition. The chosen value of k was based on three factors, first eight is divisible by four, and second it allows a sufficiently large number of fixations to ensure a reliable result, finally using an equal number of participants

	Initial fixation (2s)	Free viewing (10s)
Metric	Ambient Illuminance	Ambient Illuminance
AUC	250 1500 0	250 1500 0
EMD	1500 250 0	1500 250 0
SIM	1500 250 0	1500 250 0
CC	1500 250 0	1500 250 0
NSS	1500 250 0	250 1500 0
KLD	250 0 1500	1500 250 0

Table 7.9: Pairwise results, comparing LDR fixations similarity to HDR fixations with increasing ambient illuminance, significant at $p < 0.05$

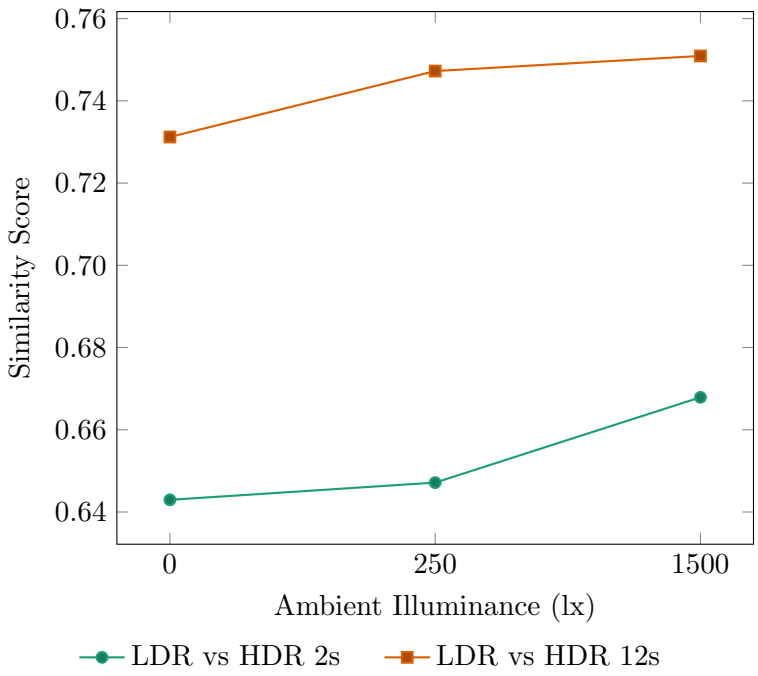


Figure 7.7: Mean Similarity scores comparing LDR vs HDR fixations for varying ambient illumination levels

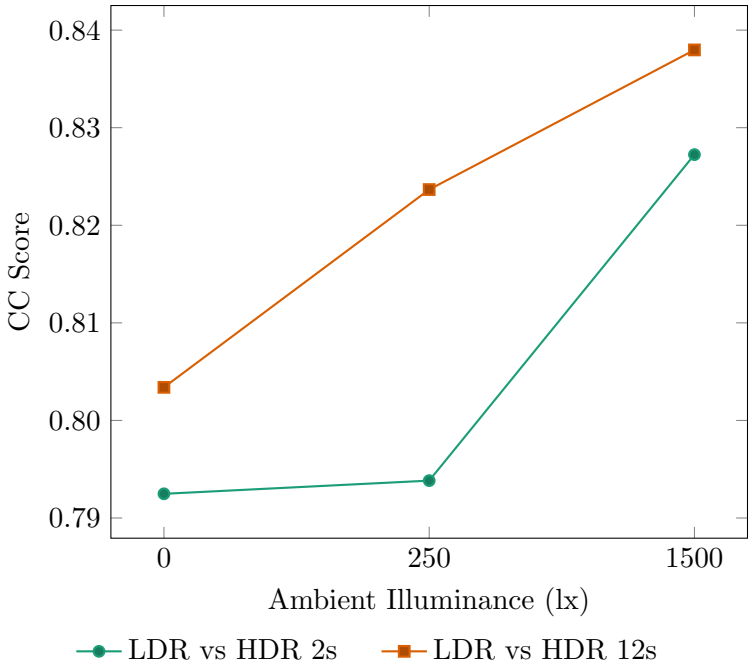


Figure 7.8: Mean CC scores comparing LDR vs HDR fixations for varying ambient illumination levels

	Ambient Light Level (lx)		
Brightness	0	250	1500
500	A	B	C
8000	D	E	F

Table 7.10: Participant groups for self prediction analysis

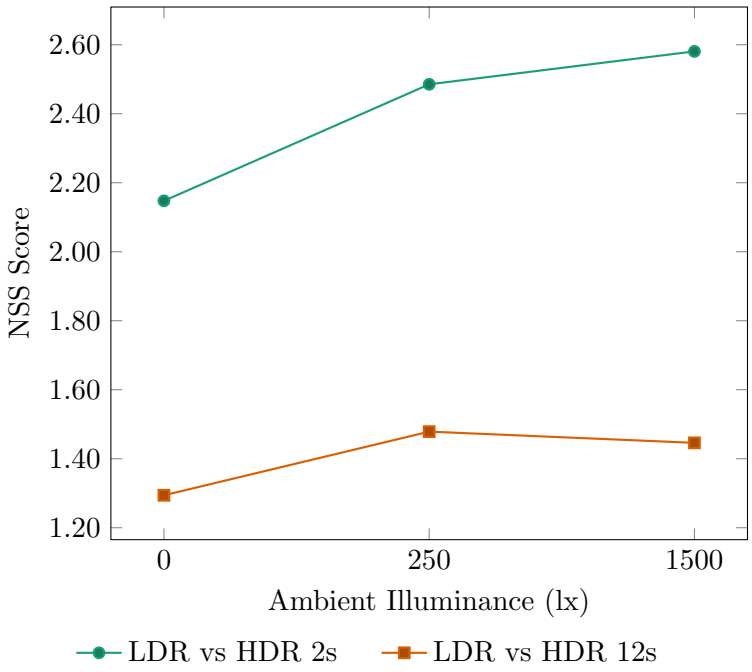


Figure 7.9: Mean NSS scores comparing LDR vs HDR fixations for varying ambient illumination levels

avoids biasing the distribution-based metrics which penalise a dissimilarity in image coverage, namely EMD and KLD. Figure 7.11, illustrates the k-fold comparison process used herein.

Analysis in the previous section observed an increase in the similarity of LDR and HDR fixations as ambient illuminance increased. The following analysis builds on this by evaluating the interaction of screen brightness and ambient illuminance and their effect on visual attention. Table 7.11 shows the results of a 2×3 factorial repeated measures ANOVA on the results from the k-fold comparisons.

Narrowing the focus to the effects of ambient illuminance on the consistency of both LDR and HDR in turn, provides a more precise idea of the effects of ambient illuminance with visual attention. Therefore, two one way ANOVAs were conducted, one for each brightness condition. The results for the LDR ANOVA can be seen in Table 7.12. Bonferroni corrected pairwise comparisons of each level can be seen in 7.13, wherein conditions grouped within coloured boxes represent comparisons where no

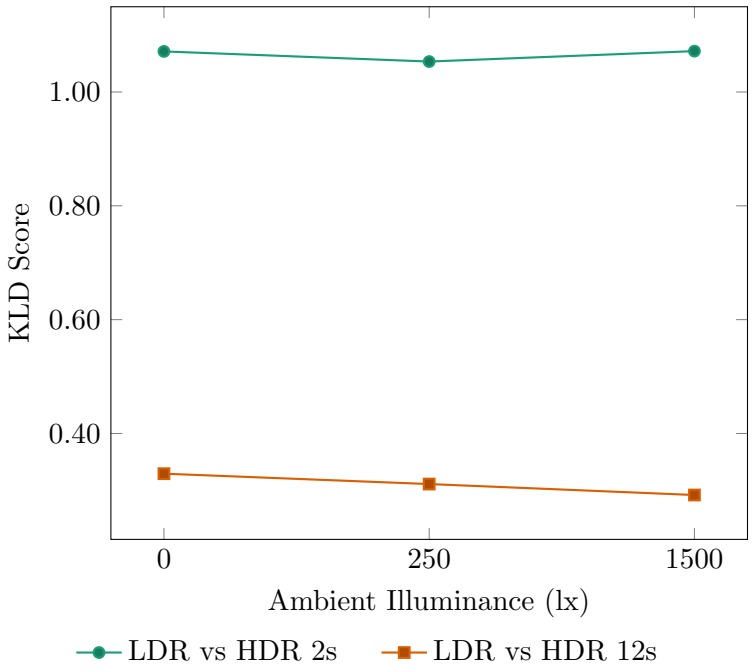


Figure 7.10: Mean KLD scores comparing LDR vs HDR fixations for varying ambient illumination levels

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	10.2	0.001	34.8	0.001
EMD	6.11	0.003	12.1	0.001
SIM	21.0	0.001	35.2	0.001
CC	20.1	0.001	26.9	0.001
NSS	25.1	0.001	32.4	0.001
KLD	8.79	0.001	21.9	0.001

Table 7.11: Results of a factorial ANOVA comparing the self-prediction consistency of fixation maps under varying brightness and ambient illuminance conditions

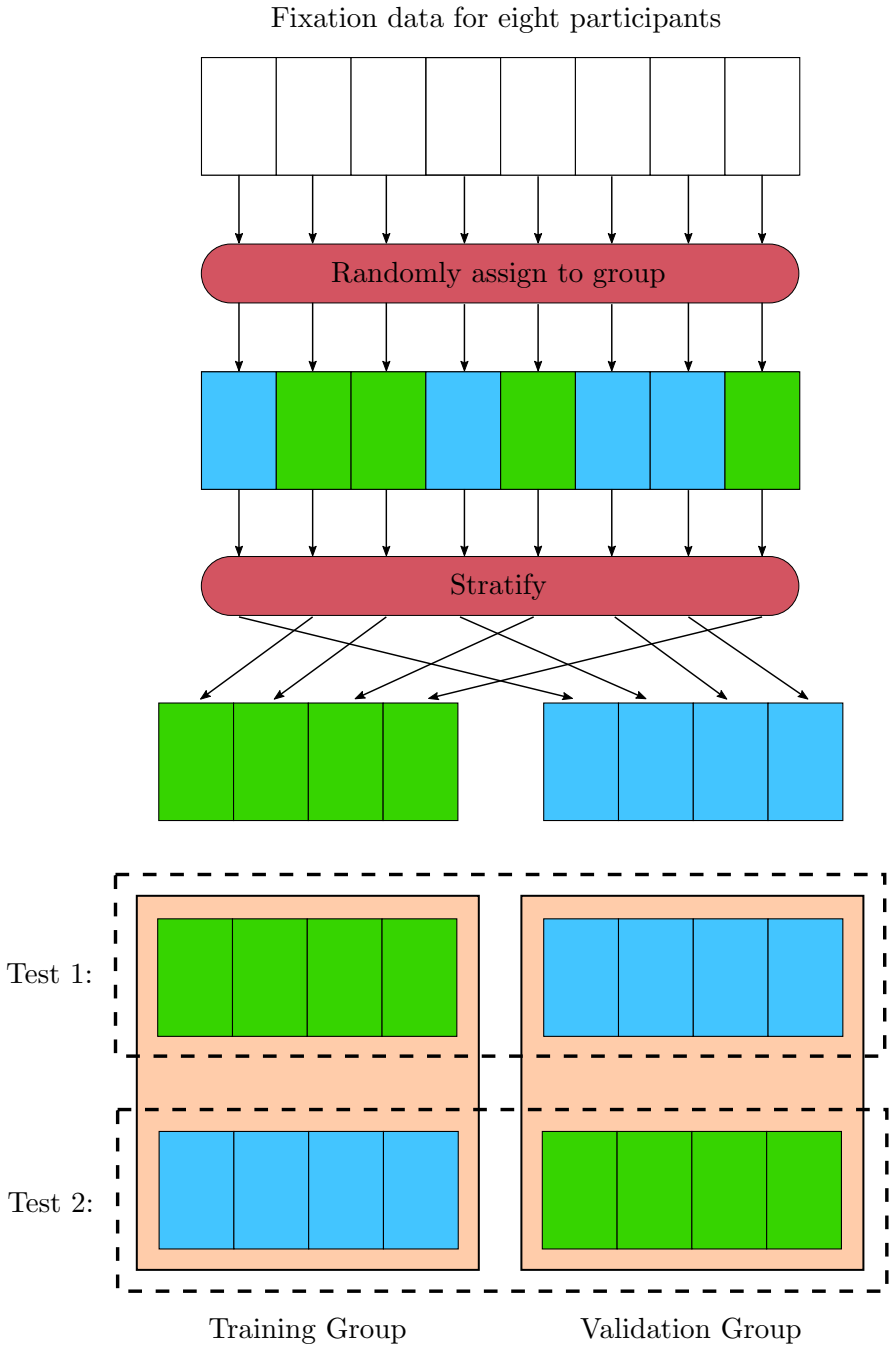


Figure 7.11: A visualisation of the k-fold algorithm used in this study

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	5.57	0.005	26.2	0.001
EMD	5.71	0.004	3.61	0.030
SIM	8.72	0.001	14.4	0.001
CC	7.46	0.001	14.7*	0.001
NSS	12.5	0.001	21.5	0.001
KLD	6.47	0.002	6.80	0.001

Table 7.12: Results of a one way ANOVA comparing the self-prediction consistency of LDR fixation maps under varying ambient illuminance conditions

significant difference, at a value of $p < 0.05$, between the populations was found. The conditions are arranged from left to right in order of decreasing mean score.

The corresponding one way factorial ANOVA for HDR conditions can be seen in 7.14, with results from the Bonferroni corrected pairwise comparisons shown in Table 7.15.

The results of the ANOVA are strong evidence that the interaction of ambient illuminance and brightness have a significant impact on the consistency, or predictability of the viewing strategies of observers. As can be seen in Table 7.11 the results of all six metrics observed significant differences between conditions, at a $p < 0.01$ level. The nature of the effect is visible in Figures 7.12, 7.13, 7.14,7.15, 7.16, and 7.17.

7.4.4 LDR as a baseline predictor under varying conditions

On review of the results presented thus far in this chapter, it is reasonable to conclude that there is some significant effect of ambient illuminance and visual attention. This in itself is interesting, however, given existing datasets [DNPN14, WS13, BJB+] are usually collected for LDR imagery in low ambient light, it is necessary to provide a benchmark for the effectiveness of such a baseline under varying conditions of ambient illuminance. Simply put; how effective is LDR eye tracking data, gathered in the dark

	Initial fixation (2s)	Free viewing (10s)
Metric	Ambient Illuminance	Ambient Illuminance
AUC	0 1500 250	0 250 1500
EMD	0 1500 250	0 250green 1500
SIM	0 1500 250	0 250 1500
CC	0 1500 250	0 250 1500
NSS	1500 0 250	0 1500 250
KLD	0 250 1500	0 250 1500

Table 7.13: Pairwise comparisons on the self-prediction consistency of LDR fixation maps under varying ambient illuminance conditions, significant at $p < 0.05$

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	6.65	0.002	15.1	0.001
EMD	1.83	0.165	10.9*	0.001
SIM	12.1	0.001	27.4	0.001
CC	14.2	0.001	16.6	0.001
NSS	27.7	0.001	15.8	0.001
KLD	2.96	0.056	21.9*	0.001

Table 7.14: Results of a one way ANOVA comparing the self-prediction consistency of HDR fixation maps under varying ambient illuminance conditions

	Initial fixation (2s)			Free viewing (10s)		
Metric	Ambient Illuminance			Ambient Illumiance		
AUC	1500	250	0	1500	250	0
EMD	1500	250	0	1500	250	0
SIM	250	1500	0	1500	250	0
CC	1500	250	0	1500	250	0
NSS	250	1500	0	250	1500	0
KLD	1500	250	0	1500	250	0

Table 7.15: Pairwise comparisons on the self-prediction consistency of HDR fixation maps under varying ambient illuminance conditions, significant at $p < 0.05$

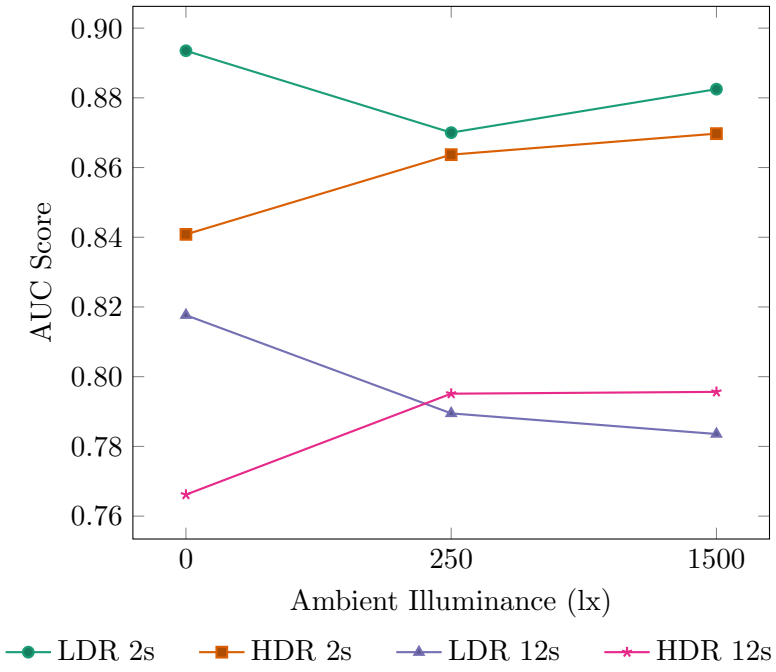


Figure 7.12: Mean AUC scores for self-prediction using k-fold cross validation

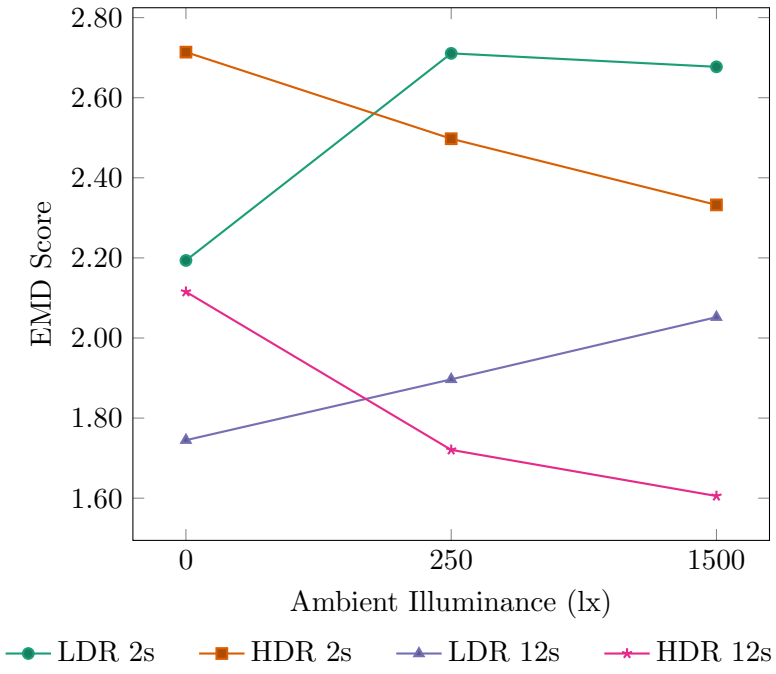


Figure 7.13: Mean EMD scores for self-prediction using k-fold cross validation

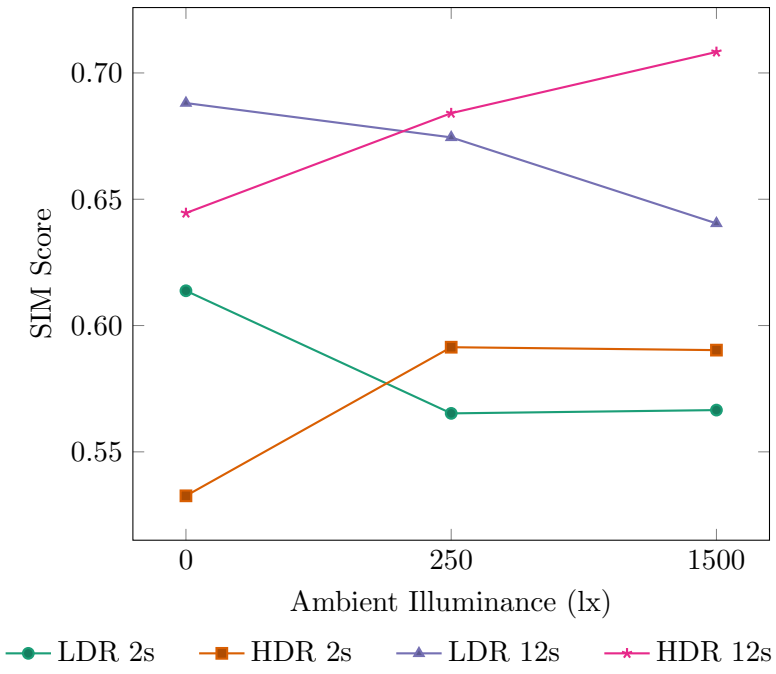


Figure 7.14: Mean Similarity scores for self-prediction using k-fold cross validation

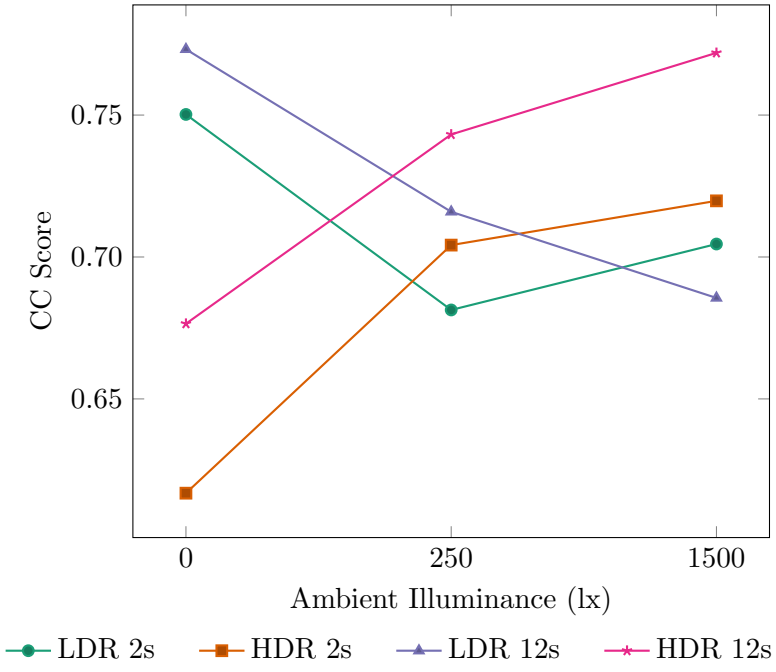


Figure 7.15: Mean CC scores for self-prediction using k-fold cross validation

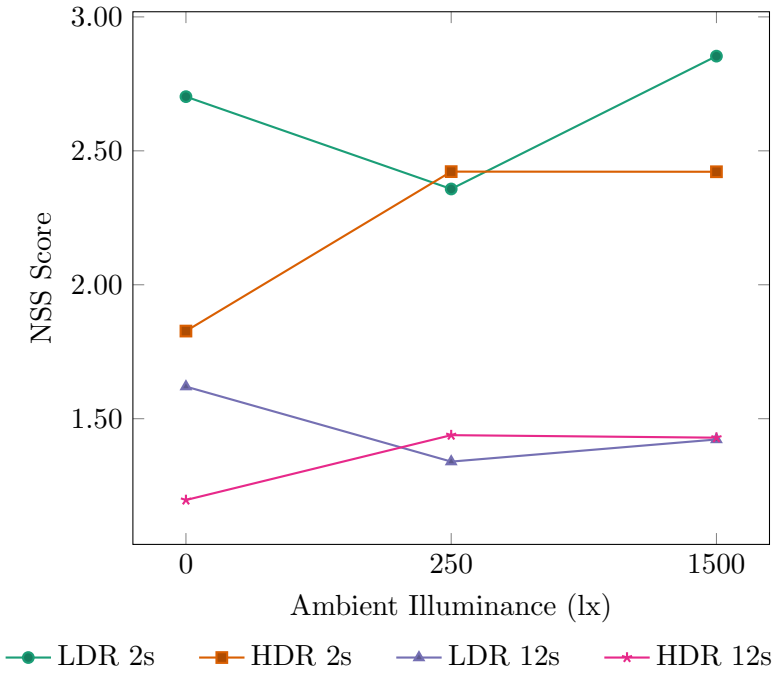


Figure 7.16: Mean NSS scores for self-prediction using k-fold cross validation

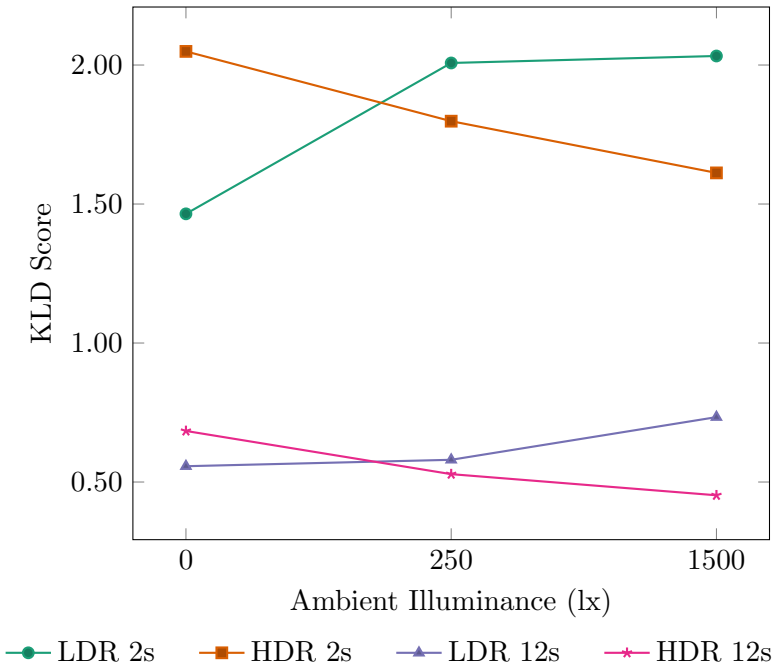


Figure 7.17: Mean KLD scores for self-prediction using k-fold cross validation

and predicting LDR and HDR content in different ambient lighting conditions?

To assess this fixation density maps, from each condition are measured against the 500 cd/m^2 and 0 lx condition using each of the six metrics. These scores are then contrasted with the 500 cd/m^2 and 0 lx condition's self-prediction score. This comparison is illustrated in Table 7.16. In order for the scores to be comparable, participant's fixations from each condition were divided into k subgroups, as in Section 7.4.3, however instead of being compared against themselves, each subset from conditions B through F is analysed against every subset of condition A. This is done to avoid biasing the results from using imbalanced numbers of participants to create the fixation density maps, which could result in a stark difference in coverage.

These results are are then averaged per condition and analysed in a 2×3 factorial ANOVA, the results of which are shown in Table 7.17.

	Ambient Light Level (lx)		
Brightness	0	250	1500
500	A	B	C
8000	D	E	F

Table 7.16: Participant groups for LDR baseline analysis

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	3.01	0.053	19.0	0.001
EMD	7.024*	0.002	11.5	0.001
SIM	22.6	0.001	33.9*	0.001
CC	18.9	0.001	34.9*	0.001
NSS	9.90	0.001	23.3*	0.001
KLD	8.90	0.001	23.4	0.001

Table 7.17: Results of inter-ambience ANOVA for all conditions, with LDR fixations, captured in low ambient light, used as a baseline predictor

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	1.67	0.193	18.4	0.001
EMD	4.98*	0.013	2.82*	0.070
SIM	8.83	0.001	8.83*	0.001
CC	4.77	0.010	9.25	0.001
NSS	8.30	0.001	15.9*	0.001
KLD	1.545	0.007	7.45*	0.002

Table 7.18: Results of inter-ambience ANOVA for LDR conditions, with LDR fixations, captured in low ambient light, used as a baseline predictor

	Initial fixation (2s)	Free viewing (10s)
Metric	Ambient Illuminance	Ambient Illuminance
AUC	1500 0 250	0 250 1500
EMD	0 250 1500	0 250 1500
SIM	0 250 1500	0 250 1500
CC	0 250 1500	0 250 1500
NSS	1500 0 250	0 1500 250
KLD	0 250 1500	250 0 1500

Table 7.19: Pairwise comparisons for LDR conditions with LDR fixations, captured in low ambient light, used as a baseline predictor, significant at $p < 0.05$

	Initial fixation (2s)		Free viewing (12s)	
Metric	F	p value	F	p value
AUC	6.59	0.002	5.10	0.008
EMD	4.19	0.017	13.5	0.001
SIM	17.5	0.001	37.0*	0.001
CC	17.4	0.001	34.4*	0.001
NSS	19.7	0.001	12.1	0.001
KLD	5.59	0.005	26.6*	0.001

Table 7.20: Results of inter-ambience ANOVA for HDR conditions, with LDR fixations, captured in low ambient light, used as a baseline predictor

	Initial fixation (2s)			Free viewing (10s)		
Metric	Ambient Illuminance			Ambient Illuminance		
AUC	1500	250	0	1500	250	0
EMD	1500	0	250	1500	250	0
SIM	1500	250	0	1500	250	0
CC	1500	250	0	1500	250	0
NSS	1500	250	0	1500	250	0
KLD	1500	250	0	1500	250	0

Table 7.21: Pairwise comparisons for HDR conditions with LDR fixations, captured in low ambient light, used as a baseline predictor, significant at $p < 0.05$

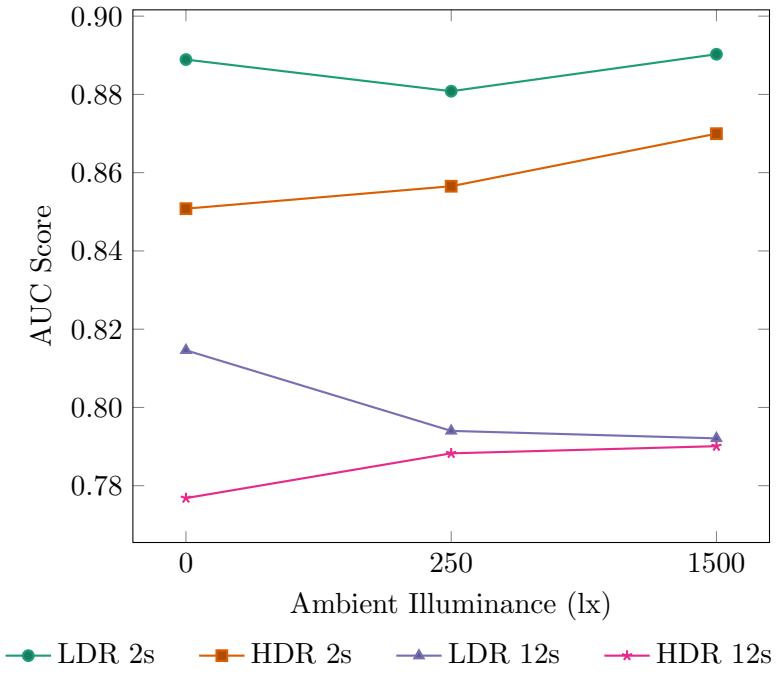


Figure 7.18: Mean AUC scores for prediction using LDR data captured in low ambient light as a baseline

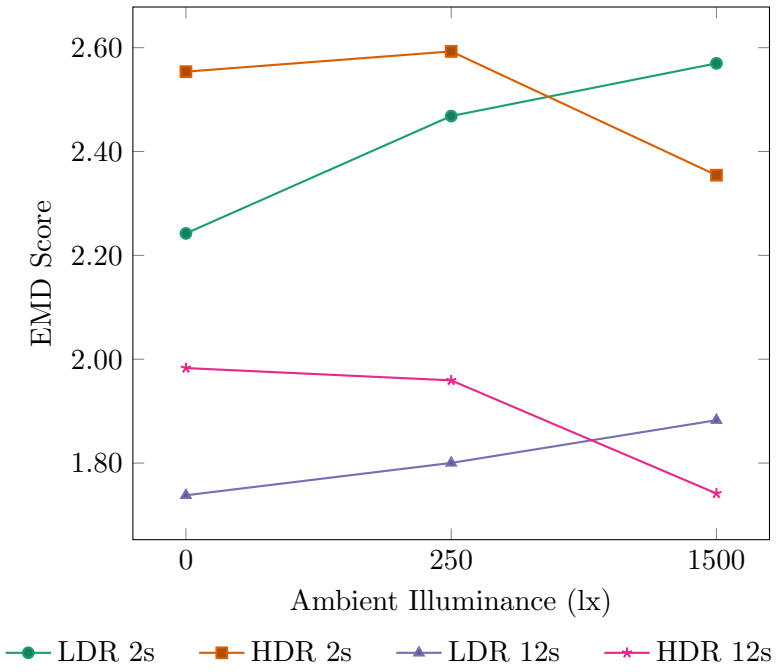


Figure 7.19: Mean EMD scores for prediction using LDR data captured in low ambient light as a baseline

7.4.5 Ambient Illumination, HDR Content and Visual Fatigue

Visual fatigue was measured using two dependent variables, the frequency at which participants blinked during the displayed image sequence, and the subjective survey, outlined in Section 7.2.3. These were gathered per participant and per group and then independently analysed in a 2×3 factorial ANOVA. The results of the analysis of the blink rate of participants found no significant effect of brightness at $p = 0.05$, ambient illuminance or the interaction of the two on blinking frequency. The recorded F scores were 0.65, 0.69 and 1.20 respectively. The trend is represented graphically in Figure 7.24, while from this graph it appears that there is an interaction between ambient illuminance and image brightness on visual fatigue, as measured by blinking frequency, this can not be concluded from the available data. This is likely due to the small number of participants per condition. It is recommended that further study be undertaken in this area however.

A 2×3 factorial repeated measures analysis of the survey questions was also

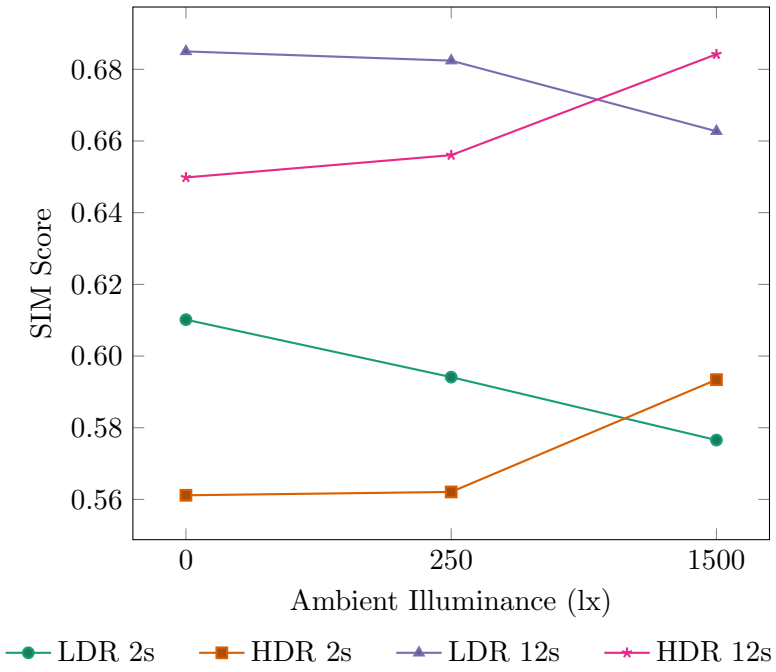


Figure 7.20: Mean Similarity scores for prediction using LDR data captured in low ambient light as a baseline

conducted. The results were gathered per participant and by condition. The results of the analysis yielded no significant results. The F scores for the effects of brightness, ambient illuminance and their interaction on visual fatigue, as measured by all questions in the survey, were 0.02, 0.8 and 1.55, respectively.

7.5 Discussion

In Section 7.2 Hypothesis **H1** stated, that the brightness of a display and the ambient illuminance of the environment will have a significant effect on visual attention and visual fatigue. In Section 7.4 k-folds cross validation was used to examine the potential interaction of these factors through a factorial ANOVA across all conditions. This allowed the assessment of the consistency of fixation data within each condition through self-prediction and its variation across the two independent variables, brightness and ambient illuminance. Table 7.11, catalogues the output of these analyses the six utilised metrics. All of which show a significant interaction for both viewing periods, this is

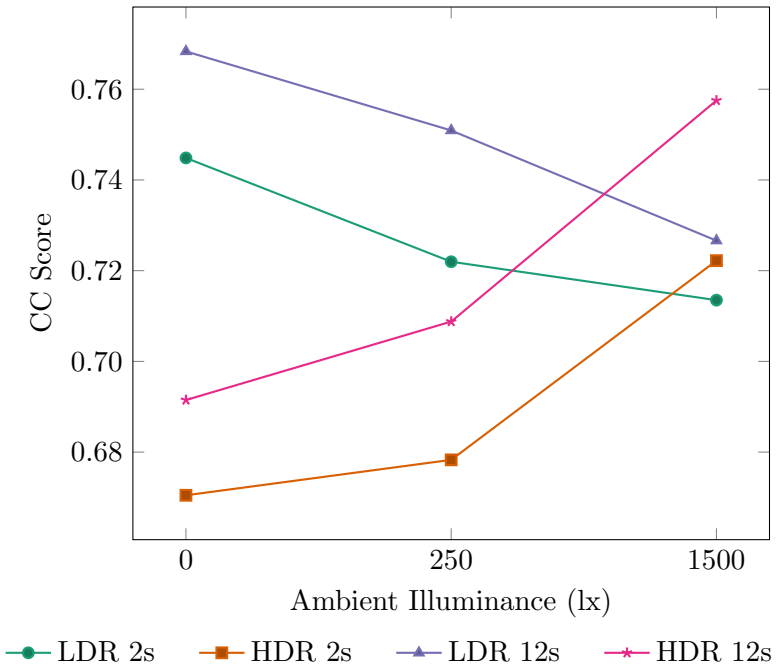


Figure 7.21: Mean CC scores for prediction using LDR data captured in low ambient light as a baseline

sufficient data for us to reject the null hypothesis, $\mathbf{H1_0}$ that there is no significant effect of ambient illuminance and brightness on visual attention. The effect is best observed on examination of Figure 7.12, 7.13, 7.14,7.15, 7.16, 7.17. These graphs show that for the majority of metrics the ability of HDR fixations to predict themselves improves as the ambient illumination increases, whereas the opposite holds true for LDR fixations.

In addition to the self-prediction analysis, the relationship of LDR and HDR fixations and the potential effect of ambient illumination was evaluated, the results of which are visible in Table 7.8, which show significant effects for most metrics across both view durations.

Results suggest that fixations under certain conditions are more predictable, rather than possessing more predictive power. While different metrics account for different types of variation within these probability maps, the outputs suggest that, even between conditions, a low variance within a population leads to a higher score, in any comparison, when attempting to predict new fixations in that condition.

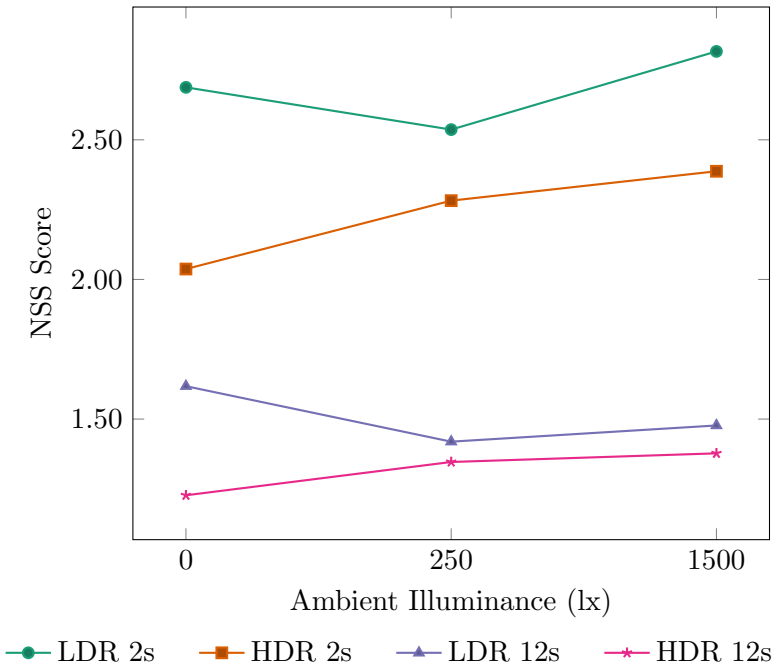


Figure 7.22: Mean NSS scores for prediction using LDR data captured in low ambient light as a baseline

In Section 7.2.1 Hypothesis $\mathbf{H2_0}$ state that the brightness of a display and the ambient illuminance of the environment will have a significant effect on visual fatigue. On review of the data examined in Section 7.4.5, it is apparent that there is insufficient data to reject the respective null hypothesis, that there is no significant effect of brightness and ambient illuminance on visual fatigue. Further study with a larger sample size is recommended to determine whether this is due to a lack of discriminative power of the study, or due to the lack of existence of an effect.

7.6 Summary

This chapter presented an eye tracking experiment, wherein ambient illuminance incident on an HDR display, and the peak brightness were manipulated to assess the viewing strategies of forty-eight participants. An analysis of the resulting data found evidence supporting the hypothesis that ambient light has a significant effect on visual attention, and that that effect is dependent, in part, on the brightness of the displayed image.

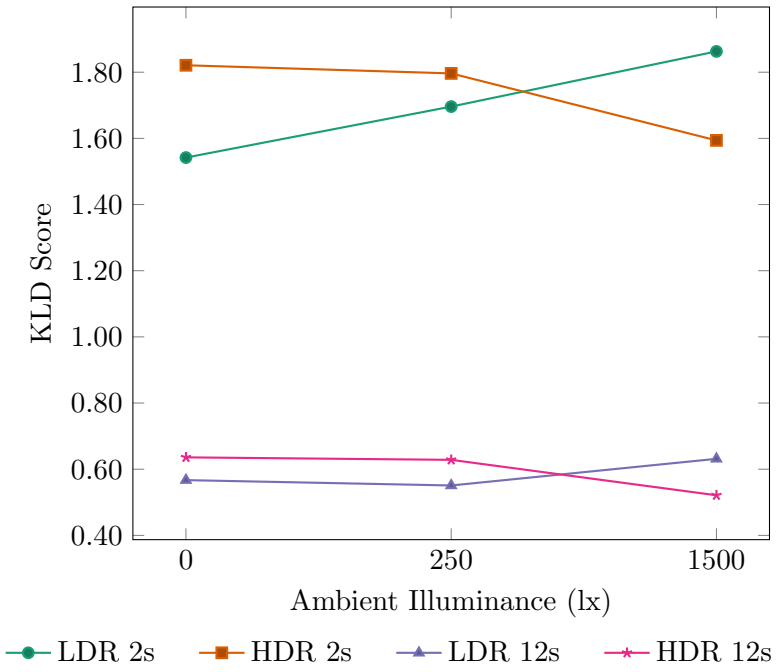


Figure 7.23: Mean KLD scores for prediction using LDR data captured in low ambient light as a baseline

This result is significant as data collection for eye-tracking studies, which are used to benchmark models of visual attention, often occurs in dark, laboratory environments. This result also has bearing on other assessments of HDR imagery, such as tone mapping operators, video compression algorithms, and subjective assessments of image quality.

An assessment of visual fatigue was also undertaken, using a subjective survey and an evaluation of the blinking frequency of participants, under the different experimental conditions. This, however, revealed no significant findings and further study is recommended.

In conclusion, the findings of this chapter suggest that while data captured in dark conditions may be effective for LDR displays, fixations recorded in brighter environments are both less representative of the population for LDR content and more consistent for HDR content. Accounting for this in future studies, may aid in the improvement of future models of visual attention for high-fidelity content. In addition, the significance of these results has implications for targeting high-fidelity rendering for cinema, which

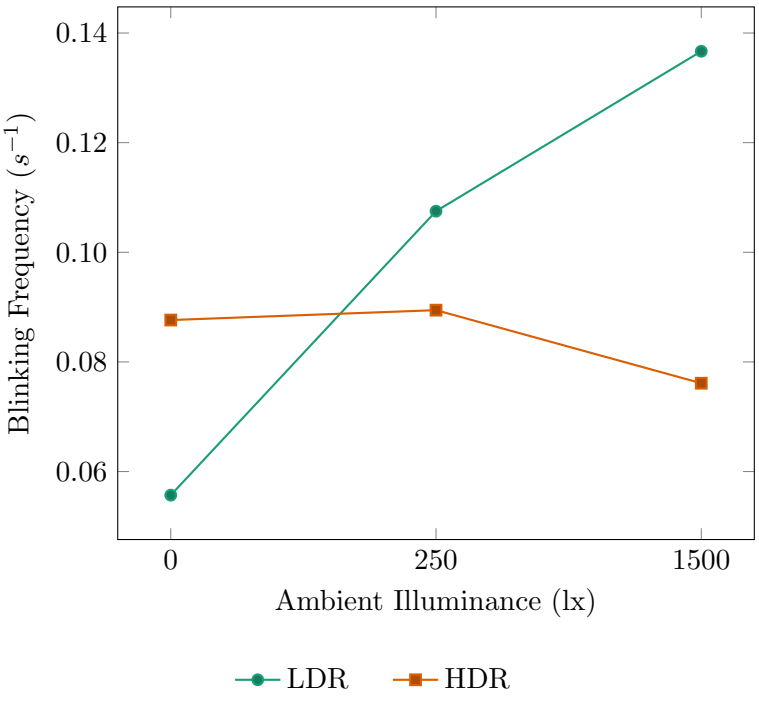


Figure 7.24: Mean blinking frequency by brightness for varying ambient illumination levels

is a traditionally dark environment, as well as stereoscopic displays, and the design of head mounted devices for augmented reality.

Chapter 8

Conclusions and Future Work

Visual attention is a complex and ever evolving field, which has many applications across a wide array of disciplines. Its adoption into high-fidelity fields such as computer graphics, HDR imaging techniques, and multi-modal virtual reality has been significant, with methods achieving improvements in perceptual quality of synthesised images and videos. Though these methods are effective, they could be enhanced through the incorporation of more targeted models of visual attention, however the underlying assumptions of visual attention in these new modalities remained untested. This work presented an analysis of these assumptions under various conditions, common to empirical analysis of both traditional and high-fidelity imaging techniques.

8.1 Selective BRDFs for High-Fidelity Rendering

First, in Chapter 5 the applicability of a conventional visual attention model [HKP06] was applied to high-fidelity rendering. The images generated as reference images for this chapter were ray-traced images with materials simulated using physically-accurate, measured materials. The experiment described in this chapter evaluated the impact of substituting analytical BRDFs in place of the aforementioned high-fidelity measured reflectance models in a selective-rendering context, using a rating experiment. The study examined the perceived similarity of images created using these reflectance methods to images created the data-driven reference model, including a hidden reference, across

three scenes. A statistical analysis of normalised ratings across all three scenes was conducted, as well as pairwise comparisons. The main effect of the method was found to be significant, at $p < 0.01$, implying that the BRDF utilised for rendering has a significant effect on visual perception. For two of the three scenes, however, pairwise comparisons found no significant differences, $p > 0.05$, between the rating of the hidden reference image and the selectively rendered images. For all three scenes the selectively rendered images were rated higher than those rendered with their respective analytic BRDFs. All the selectively rendered methods provided modest reductions in computational cost, though salient diffuse demonstrated the largest mean computational saving (10.8%). This suggests that visual saliency can be used to improve computational performance by replacing computationally expensive materials with cheaper analytic models in less visually important regions of a scene. This builds on existing work in applying visual attention to high-fidelity rendering, adding an additional manner in which to alleviate the computational burden of generating realistic images, which could be used in conjunction with existing selective rendering techniques and visual attention models.

8.2 HDR Eye-tracking

Chapter 6 evaluated visual attention in the context of high dynamic range imaging, an high-fidelity imaging research that has received much attention in recent years. HDR imaging is implicit in physically-based rendering systems, which attempt to approximate light transport as accurately as is feasible, before transforming the rendered frame to a target luminance for display conventional LDR screen. Therefore it is important for both the display of real world and virtual HDR scenes to be able to target visual attention to a particular luminance.

This chapter therefore examined the effects of HDR luminance levels on visual attention by considering the viewing strategies of 48 participants utilising 80 HDR images from the Fairchild HDR photographic survey [Fai07], scaled to four luminance levels, 500, 1,000, 2,000 and 4,000 $\mathrm{cd/m^2}$. To the best of the author's knowledge this is the first work to evaluate visual attention the effects of controlled HDR luminance

levels on visual attention.

The results of the statistical analysis of the created eye-tracking dataset revealed significant effects of screen brightness on the viewing strategies of observers. Performing contrasts across conditions found a significant effect, at $p < 0.05$, of brightness on the similarity of fixation groups to an LDR baseline, for the majority of metrics. The observed trend showed a decrease in similarity as screen brightness increased. The main effect of an ANOVA of the self-prediction or uniformity of fixations between brightness levels, was also found to be significant, at $p < 0.05$ for the majority of metrics.

This is an important finding, as it reveals dynamic range needs to be considered when creating new visual attention models. In addition it informs the creation of new HDR datasets, as from the data it is apparent that datasets captured in laboratory conditions are less uniform than their LDR counterparts, which suggests that larger, more relevant, datasets may be needed for HDR content.

8.3 Ambient Light, Visual Attention and Visual Fatigue

Chapter 7 built on the work of 6 and incorporated an analysis of the effect of environmental illumination on viewing strategies and visual fatigue for both LDR and HDR content. A second eye-tracking dataset was created, containing three ambient light conditions, of 0, 250 and 1,500 lx. Analysis revealed that ambient light has a significant effect on visual attention, which varies significantly with the brightness of the displayed image. The effects of ambient light were found to be insignificant on visual fatigue, which was evaluated both through a subjective rating survey and a quantitative analysis of blink rates of participants. A qualitative examination of the resulting graphs suggests that an effect existed, however there was not enough power to demonstrate whether a true effect existed, with power reported at < 0.2 the analysis of survey questions and < 0.3 in blink rate analysis.

Turning to visual attention, an analysis of LDR as a baseline predictor under varying ambient light levels found that fixations captured for LDR content were most similar to those for HDR content when the former was captured in low ambient light and the latter in bright ambient light. This is potentially caused by difficulties adapting to

the entire luminance range, though further study is needed. The work also found a difference in the uniformity of fixations across ambient brightness levels, for both LDR and HDR content. Results showed that as ambient light increased, variation within the population rose for LDR content and fell for HDR content. This is likely due to the variation in detail visible in both conditions as a result of the ambient lighting.

The implications of this study, the first analysis of visual attention under varying environmental illumination for both LDR and HDR content, are potentially far reaching. The observed difference in viewing strategies was significant to $p < 0.01$ in the majority of cases for six metrics of saliency map comparison. This should inform not only eye tracking studies, and the benchmarking of visual attention models, but the conditions in which subjective and objective tests are conducted for high-fidelity imagery.

8.4 Contributions

This thesis has built on the existing knowledge of visual attention for high-fidelity imaging, expanding the techniques for exploiting and the empirical basis for the analysing of visual attention models in fidelity-contexts. In Chapter 1 the following research question was posed:

> *How can in-depth knowledge of visual attention be used effectively when generating high-fidelity imagery?*

The primary contribution of this thesis is an analysis of this question as it pertains to both selective rendering, wherein a method was proposed and tested which exploits visual attention for generating high-fidelity rendered images, and HDR content and subjective evaluation of high-fidelity imaging techniques in environments with controlled ambient light conditions, through the quantitative analysis of how visual attention is affected by these factors. The primary contributions of this thesis are as follows:

1. An algorithm for alleviating the cost of physically-based ray tracing by exploiting visual attention to reduce the cost of reflectance simulation in virtual environments.

2. A subjective rating experiment demonstrating the viability of the above method.

3. An eye-tracking dataset, containing the fixations of 48 participants for 80 HDR images, representing the largest HDR eye-tracking dataset collected to date.

4. An analysis of the effects of screen brightness on the consistency of visual attention.

5. An analysis of the viability of using LDR eye tracking datasets to predict viewing strategies of HDR images.

6. The first eye-tracking dataset of LDR and HDR images to be captured under varying ambient illuminance.

7. Eye tracking data captured for HDR images displayed on a screen capable of outputting luminance values of 8,000 $\mathrm{cd/m^2}$. The highest recorded experimentally.

8. An analysis of ambient illuminance on visual attention for LDR and HDR content.

8.5 Limitations and Future Work

This section suggests future work that could build on the research presented in this thesis and discusses the limitations of the work presented herein.

Selective BRDFs

The method described in Chapter 5 is dependent on a precomputed saliency map. Thus computational savings are mitigated by the time necessary to generate a sufficiently accurate image to create a reliable saliency map. This limitation could be partially mitigated by expanding the presented method to take radiance values directly from the frame-buffer in the renderer, though this would still introduce a computational cost. This work could be expanded to approximate high quality BRDFs that incorporate physical approximations of diffraction [LKYU12][HP15] and the polarisation of incident light [HTSG91]. These functions are more mathematically complicated and in some cases [HP15] require pre-computation of the geometric shadowing and masking functions. The algorithm for adjusting the proportion of samples distributed to each BRDF should be expanded. A naive linear approach was taken to distribution, however the method could benefit from introducing thresholds and non-linear scaling.

HDR Eye tracking and Ambient Illumination

The data collected in Chapter 6 and 7 came primarily from University science students. This could have a potential biasing effect on the results. The eye tracking datasets provided sufficient data for analysis of the underlying variables, but there are not sufficient images contained for the training of a reliable machine learning model. This could be expanded through the analysis of a larger number of participants and images. Though few datasets of HDR images exist, so these would need to be created beforehand.

The existing raw data however has much value to further research and can be reprocessed to examine the dwell maps of participants, these maps account for the length of fixations. An analysis of this could advance the understanding of the manner in which fixations on HDR content diverge from those on LDR content.

This results gathered in Chapter 7 should be used in conjunction with the results of Chapter 6 to inform a new saliency model, for high-fidelity imaging in targeted environments. This model should then be used in conjunction with the method presented in Chapter 5 to reassess the effects of adaptively adjusting reflection models for high-fidelity rendered images.

A limitation of the both Chapter 6 and 7 is that the scaling method employed can desaturate colour channels at high brightnesses. This needs to be taken into account utilising the dataset for the evaluation of HDR visual attention model or different high-fidelity techniques. Capturing HDR image data in calibrated environments could mitigate this but would be costly to collect in large numbers.

8.6 Final Remarks

Visual attention is an ever advancing field, with new models being proposed for stereo imaging, HDR content, and virtual reality. However, the understanding of underlying features introduced by these new display modalities has remained underdeveloped. Identifying these new features is important, as an intuitive understanding of the process by which attention is directed by an observer can only serve to enhance existing models.

This thesis has investigated the use of visual attention for high-fidelity imaging,

and while not solving this problem, has hopefully provided a firm foundation on which future work can build.

Bibliography

[AF34] Francis Heed Adler and Maurice Fliegelman. Influence of fixation on the visual acuity. *Archives of Ophthalmology*, 12(4):475–483, 1934.

[AHES09] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, pages 1597–1604. IEEE, 2009.

[AS00] Michael Ashikhmin and Peter Shirley. An anisotropic phong brdf model. *Journal of graphics tools*, 5(2):25–32, 2000.

[Ass03] ITU Radiocommunication Assembly. *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, 2003.

[BADC17] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced high dynamic range imaging*. CRC press, 2017.

[Bar52] Horace B Barlow. Eye movements during fixation. *The Journal of Physiology*, 116(3):290–306, 1952.

[BCDZ+14] Simone Benedetto, Andrea Carbone, Véronique Drai-Zerbib, Marco Pedrotti, and Thierry Baccino. Effects of luminance and illuminance on visual fatigue and arousal during digital reading. *Computers in Human Behavior*, 41:112–119, 2014.

[BCP03] Ross Brown, Luke Cooper, and Binh Pham. Visual attention-based polygon level of detail management. In *Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 55–ff. ACM, 2003.

[BDBR+16] Tim Bradley, Kurt Debattista, Thomas Bashford-Rogers, Carlo Harvey, E Doukakis, and Alan Chalmers. Selective brdfs for high fidelity rendering. In *Proceedings of the conferece on Computer Graphics & Visual Computing*, pages 57–64. Eurographics Association, 2016.

[BDDPN15] Amin Banitalebi-Dehkordi, Yuanyuan Dong, Mahsa T Pourazad, and Panos Nasiopoulos. A learning-based visual saliency fusion model for high dynamic range video (lbvs-hdr). In *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pages 1541–1545. IEEE, 2015.

[BDPN16] Amin Banitalebi-Dehkordi, Mahsa T Pourazad, and Panos Nasiopoulos. A learning-based visual saliency prediction model for stereoscopic 3d video (lbvs-3d). *Multimedia Tools and Applications*, pages 1–32, 2016.

[BI13] Ali Borji and Laurent Itti. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207, 2013.

[BJB+] Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. Mit saliency benchmark.

[BJO+16] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. What do different evaluation metrics tell us about saliency models? *arXiv preprint arXiv:1604.03605*, 2016.

[Bli77] James F Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH Computer Graphics*, volume 11, pages 192–198. ACM, 1977.

[BLPW14] Adam Brady, Jason Lawrence, Pieter Peers, and Westley Weimer. genbrdf: Discovering new analytic brdfs with genetic programming, 2014.

[BPT10] Roland Brémond, Josselin Petit, and Jean-Philippe Tarel. Saliency maps of high dynamic range images. In *European Conference on Computer Vision*, pages 118–130. Springer, 2010.

[Bra97] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[BRT08] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. It's in your eyes: towards context-awareness and mobile hci using wearable eog goggles. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 84–93. ACM, 2008.

[BRY98] Andrew Barron, Jorma Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2760, 1998.

[BSH12] Mahdi M Bagher, Cyril Soler, and Nicolas Holzschuch. Accurate fitting of measured reflectances using a shifted gamma micro-facet distribution. In *Computer Graphics Forum*, volume 31, pages 1509–1518. Wiley Online Library, 2012.

[BVK+05] Clayton Blehm, Seema Vishnu, Ashbala Khattak, Shrabanee Mitra, and Richard W. Yee. Computer vision syndrome: A review. *Survey of Ophthalmology*, 50(3):253 – 262, 2005.

[CARR14] Lounis Chermak, Nabil Aouf, Mark Richardson, and Magnus Ramage. Hdr imaging for feature tracking in challenging visibility scenes. *Kybernetes*, 43(8), 2014.

[CC73] Tom N Cornsweet and Hewitt D Crane. Accurate two-dimensional eye tracker using first and fourth purkinje images. *JOSA*, 63(8):921–928, 1973.

[CCL02] Kirsten Cater, Alan Chalmers, and Patrick Ledda. Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 17–24. ACM, 2002.

[CCW03] Kirsten Cater, Alan Chalmers, and Greg Ward. Detail to attention: exploiting visual tasks for selective rendering. In *ACM International Conference Proceeding Series*, volume 44, pages 270–280, 2003.

[CDMPdS07] Alan Chalmers, Kurt Debattista, Georgia Mastoropoulou, and Luis Paulo dos Santos. There-reality: selective rendering in high fidelity virtual environments. *The International Journal of Virtual Reality*, 6(1):1–10, 2007.

[Cos94] MA Costanza. Visual and ocular symptoms related to the use of video display terminals. *J Behav Optom*, 5:31–36, 1994.

[CT82] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24, 1982.

[DG52] RW Ditchburn and BL Ginsborg. Vision with a stabilized retinal image. *Nature*, 170(4314):36–37, 1952.

[DHKS86] H Doi, Y Hara, Y Kenbo, and M Shiba. Image sensor. *Japanese Patent*, pages 08–223491, 1986.

[DM08] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM, 2008.

[DNPN14] Yuanyuan Dong, Eleni Nasiopoulos, Mahsa T Pourazad, and Panos Nasiopoulos. High dynamic range video eye tracking dataset. In *Proc. 2nd Int. Conf. Electron., Signal Process. Commun.*, pages 56–59, 2014.

[Doh25] Gösta Dohlman. *Physikalische und physiologische Studien zur Theorie des kalorischen Nystagmus...*, volume 5. Almquist & Wiksells boktryckeri, 1925.

[DPN16] Yuanyuan Dong, Mahsa T Pourazad, and Panos Nasiopoulos. Human visual system-based saliency detection for high dynamic range content. *IEEE Transactions on Multimedia*, 18(4):549–562, 2016.

[Duc02] Andrew T Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, 2002.

[DVGNK97] K Dana, Bram Van Ginneken, S Nayar, and J Koenderink. Columbia-utrecht reflectance and texture database, 1997.

[ELW+13] Ulrich Engelke, Hantao Liu, Junle Wang, Patrick Le Callet, Ingrid Heynderickx, Hans-Jürgen Zepernick, and Anthony Maeder. Comparative study of fixation density maps. *IEEE Transactions on Image Processing*, 22(3):1121–1133, 2013.

[EMZ09] Ulrich Engelke, Anthony Maeder, and Hans-Jurgen Zepernick. Visual attention modelling for subjective image quality databases. In *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*, pages 1–6. IEEE, 2009.

[ENSB13] Christian Eisenacher, Gregory Nichols, Andrew Selle, and Brent Burley. Sorted deferred shading for production path tracing. In *Computer Graphics Forum*, volume 32, pages 125–132. Wiley Online Library, 2013.

[Fai07] Mark D Fairchild. The hdr photographic survey. In *Color and Imaging Conference*, volume 2007, pages 233–238. Society for Imaging Science and Technology, 2007.

[Faw06] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[FGE+14] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In *Digital Photography*, page 90230X, 2014.

[FLL+17] Yuming Fang, Jianjun Lei, Jia Li, Long Xu, Weisi Lin, and Patrick Le Callet. Learning visual saliency from human fixations for stereoscopic images. *Neurocomputing*, 2017.

[FRC10] Simone Frintrop, Erich Rome, and Henrik I Christensen. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)*, 7(1):6, 2010.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[GDS14] Steven Galea, Kurt Debattista, and Sandro Spina. Gpu-based selective sparse sampling for interactive high-fidelity rendering. In *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2014 6th International Conference on*, pages 1–8. IEEE, 2014.

[GTL23] Edgar J George, Julius A Toren, and Jay Webb Lowell. Study of the ocular movements in the horizontal plane. *American Journal of Ophthalmology*, 6(10):833–838, 1923.

[Har11] Carlo Harvey. *Modality based perception for selective rendering*. PhD thesis, University of Warwick, 2011.

[HE12] Christopher G Healey and James T Enns. Attention and visual memory in visualization and computer graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 18(7):1170–1188, 2012.

[HHKR88] H Heuer, G Hollendiek, H Kröger, and T Römer. Rest position of the eyes and its effect on viewing distance and visual fatigue in computer display work. *Zeitschrift fur experimentelle und angewandte Psychologie*, 36(4):538–566, 1988.

[HKP06] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2006.

[HM82] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[HP15] Nicolas Holzschuch and Romain Pacanowski. *A physically accurate reflectance model combining reflection and diffraction.* PhD thesis, INRIA, 2015.

[HTSG91] Xiao D He, Kenneth E Torrance, Francois X Sillion, and Donald P Greenberg. A comprehensive physical model for light reflection. In *ACM SIGGRAPH computer graphics*, volume 25, pages 175–186. ACM, 1991.

[Hue08] Edmund Burke Huey. *The psychology and pedagogy of reading.* The Macmillan Company, 1908.

[IK00] Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10):1489–1506, 2000.

[IKK+09] Takuya Iwanami, Ayano Kikuchi, Takashi Kaneko, Keita Hirai, Natsumi Yano, Toshiya Nakaguchi, Norimichi Tsumura, Yasuhiro Yoshida, and Yoichi Miyake. The relationship between ambient illumination and

psychological factors in viewing of display images. In *Color Imaging: Displaying, Processing, Hardcopy, and Applications*, page 72410L, 2009.

[IKN98] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (11):1254–1259, 1998.

[Ins17] SensoMotoric Instruments. Smi eye tracking from science to your application. `https://www.smivision.com/`, January 2017.

[IR12] ITU-R. Bt.500-13 "methodology for the subjective assessment of the quality of television pictures". *International Telecommunications Union, Geneva*, 1 2012.

[Jam13] William James. *The principles of psychology*. Read Books Ltd, 2013.

[Jav78] Emile Javal. Essai sur la physiologie de la lecture. *Annales d'Ocilistique*, 80:61–73, 1878.

[JDT12] Tilke Judd, Frédo Durand, and Antonio Torralba. A benchmark of computational models of saliency to predict human fixations. In *MIT Technical Report*, 2012.

[JEDT09] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113. IEEE, 2009.

[JHDZ15] Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. Salicon: Saliency in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1072–1080, 2015.

[JMLH01] Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518. ACM, 2001.

[Kaj86] James T Kajiya. The rendering equation. In *ACM Siggraph Computer Graphics*, volume 20, pages 143–150. ACM, 1986.

[KDCM14] George Alex Koulieris, George Drettakis, Douglas Cunningham, and Katerina Mania. C-lod: Context-aware material level-of-detail applied to mobile graphics. In *Computer Graphics Forum*, volume 33, pages 41–49. Wiley Online Library, 2014.

[KMS+06] Kristin Koch, Judith McLean, Ronen Segev, Michael A Freed, Michael J Berry, Vijay Balasubramanian, and Peter Sterling. How much the eye tells the brain. *Current Biology*, 16(14):1428–1434, 2006.

[KMZS07] Grzegorz Krawczyk, Rafal Mantiuk, Dorota Zdrojewska, and Hans-Peter Seidel. Brightness adjustment for hdr and tone mapped images. In *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on*, pages 373–381. IEEE, 2007.

[KS39] Maurice G Kendall and B Babington Smith. The problem of m rankings. *The annals of mathematical statistics*, 10(3):275–287, 1939.

[KS01] Kenichi Kaneko and Kazuyoshi Sakamoto. Spontaneous blinks as a criterion of visual fatigue during prolonged work on visual display terminals. *Perceptual and Motor Skills*, 92(1):234–250, 2001.

[KTB14] Matthias Kümmerer, Lucas Theis, and Matthias Bethge. Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet. *arXiv preprint arXiv:1411.1045*, 2014.

[KU87] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

[Kul97] Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1997.

[KVDS96] Jan J Koenderink, Andrea J Van Doorn, and Marigo Stavridi. Bidirectional reflection distribution function expressed in terms of surface scattering modes. In *European Conference on Computer Vision*, pages 28–39. Springer, 1996.

[LDC06] Peter Longhurst, Kurt Debattista, and Alan Chalmers. A gpu based saliency map for high-fidelity selective rendering. In *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pages 21–29. ACM, 2006.

[Lew94] Robert R Lewis. Making shaders more physically plausible. In *Computer Graphics Forum*, volume 13, pages 109–120. Wiley Online Library, 1994.

[LFH15] Ajay Luthra, Edouard Franois, and Walt Husak. Call for evidence (cfe) for hdr and wcg video coding. *ISO/IEC JTC1/SC29/WG11 MPEG, Geneva, Switzerland, Tech. Rep N*, 15083, 2 2015.

[LFI+10] Marc Lambooij, Marten Fortuin, Wijnand Ijsselsteijn, Bruce Evans, and Ingrid Heynderickx. Measuring visual fatigue and visual discomfort associated with 3-d displays. *Journal of the Society for Information Display*, 18(11):931–943, 2010.

[LFTG97] Eric PF Lafortune, Sing-Choong Foo, Kenneth E Torrance, and Donald P Greenberg. Non-linear approximation of reflectance functions. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 117–126. ACM Press/Addison-Wesley Publishing Co., 1997.

[LHP10] Eui Chul Lee, Hwan Heo, and Kang Ryoung Park. The comparative measurements of eyestrain caused by 2d and 3d displays. *IEEE Transactions on Consumer Electronics*, 56(3), 2010.

[LKYU12] Joakim Löw, Joel Kronander, Anders Ynnerman, and Jonas Unger. Brdf models for accurate and efficient rendering of glossy surfaces. *ACM Transactions on Graphics (TOG)*, 31(1):9, 2012.

[LLF13] Sébastien Lasserre, Fabrice LeLéannec, and Edouard Francois. Description of hdr sequences proposed by technicolor. *ISO/IEC JTC1/SC29/WG11 JCTVC-P0228], IEEE, San Jose, USA*, 2013.

[Low04] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[LPWM09] Eui Chul Lee, Kang Ryoung Park, Mincheol Whang, and Kyungha Min. Measuring the degree of eyestrain caused by watching lcd and pdp devices. *International Journal of Industrial Ergonomics*, 39(5):798–806, 2009.

[Mat74] Ethel Matin. Saccadic suppression: a review and an analysis. *Psychological bulletin*, 81(12):899, 1974.

[MBB+15] Miguel Melo, Maximino Bessa, Luís Barbosa, Kurt Debattista, and Alan Chalmers. Screen reflections impact on hdr video tone mapping for mobile devices: an evaluation study. *EURASIP Journal on Image and Video Processing*, 2015(1):44, 12 2015.

[Mca02] David Kirk Mcallister. *A generalized surface appearance representation for computer graphics*. PhD thesis, University of North Carolina at Chapel Hill, 2002.

[McG17] Morgan McGuire. Computer graphics archive, July 2017. `https://casual-effects.com/data`.

[MM05] Carlos H Morimoto and Marcio RM Mimica. Eye gaze tracking techniques for interactive applications. *Computer vision and image understanding*, 98(1):4–24, 2005.

[Mor11] Geoffrey Morrison. Contrast ratio (or how every tv manufacturer lies to you). Online, May 2011.

[MP94] Steve Mann and Rosalind Picard. *Being undigital with digital cameras*. MIT Media Lab Perceptual, 1994.

[MPBM03] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, July 2003.

[MT11] Eugen Marx and Wilhelm Trendelenburg. Über die genauigkeit der einstellung des auges beim fixieren. *Zeitschrift für Sinnesphysiologie*, 45:87–102, 1911.

[MTM12] Rafał K Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. In *Computer Graphics Forum*, volume 31, pages 2478–2491. Wiley Online Library, 2012.

[Muk17] Ratnajit Mukherjee. *Accurate light and colour reproduction in high dynamic range video compression.* PhD thesis, University of Warwick, 2017.

[MYS95] F Malbet, JW Yu, and M Shao. High-dynamic-range imaging using a deformable mirror for space coronography. *Publications of the Astronomical Society of the Pacific*, pages 386–398, 1995.

[Nac59] Jacob Nachmias. Two-dimensional motion of the retinal image during monocular fixation. *JOSA*, 49(9):901–908, 1959.

[NDK14] Eleni Nasiopoulos, Yuanyuan Dong, and Alan Kingstone. Evaluation of high dynamic range content viewing experience using eye-tracking data. In *Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference on*, pages 13–17. IEEE, 2014.

[NDM05] Addy Ngan, Frédo Durand, and Wojciech Matusik. Experimental analysis of brdf models. *Rendering Techniques*, 2005(16th):2, 2005.

[NDSLCP14] Manish Narwaria, Matthieu Perreira Da Silva, Patrick Le Callet, and Romuald Pepion. Tone mapping based hdr compression: Does it affect visual experience? *Signal Processing: Image Communication*, 29(2):257–273, 2014.

[Nev69] John A Nevin. Signal detection theory and operant behavior: A review of david m. green and john a. swets' signal detection theory and psychophysics. *Journal of the Experimental Analysis of Behavior*, 12(3):475–480, 1969.

[New28] SM Newhall. Instrument for observing ocular movements. *The American Journal of Psychology*, 40(4):628–629, 1928.

[NKHE15] Hiromi Nemoto, Pavel Korshunov, Philippe Hanhart, and Touradj Ebrahimi. Visual attention in ldr and hdr images. In *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, number EPFL-CONF-203873, 2015.

[NM00] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 472–479. IEEE, 2000.

[Not05] Hans-Christoph Nothdurft. Salience of feature contrast. 2005.

[ON95] Michael Oren and Shree K Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995.

[PBT09] Josselin Petit, Roland Brémond, and Jean-Philippe Tarel. Saliency maps of high dynamic range images. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, pages 134–134. ACM, 2009.

[Pec34] Robert H Peckham. Foveal projection during ductions. *Archives of Ophthalmology*, 12(4):562–566, 1934.

[Pho75] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[PIIK05] Robert J Peters, Asha Iyer, Laurent Itti, and Christof Koch. Components of bottom-up gaze allocation in natural images. *Vision research*, 45(18):2397–2416, 2005.

[Poi17] PointGrey. Flea2 (1394b) cameras for industrial, life science, traffic and security applications, 2017.

[Pos80] Michael I Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980.

[psn13] Peak signal-to-noise ratio as an image quality metric. white paper, September 2013.

[PW08] Ofir Pele and Michael Werman. A linear time histogram metric for improved sift matching. *Computer Vision–ECCV 2008*, pages 495–508, 2008.

[PW09] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *Computer vision, 2009 IEEE 12th international conference on*, pages 460–467. IEEE, 2009.

[Ray98] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[Rec12] ITURBT Recommendation. 2022:"general viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays". *ITU Radiocommunication*, 2012.

[RGLC17] Yashas Rai, Jesús Gutiérrez, and Patrick Le Callet. A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 205–210. ACM, 2017.

[RHD+10] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[RR50] Floyd Ratliff and Lorrin A Riggs. Involuntary motions of the eye during monocular fixation. *Journal of experimental psychology*, 40(6):687, 1950.

[RTG00] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[Sai95] Kuniaki Saito. Electronic image pickup device. *Japanese Patent*, pages 07–254965, 1995.

[SB91] Michael J Swain and Dana H Ballard. Color indexing. *International journal of computer vision*, 7(1):11–32, 1991.

[SCCD04] Veronica Sundstedt, Alan Chalmers, Kirsten Cater, and Kurt Debattista. Top-down visual attention for efficient rendering of task related scenes. In *VMV*, pages 209–216, 2004.

[Sch22] Eduard Schott. Uber die registrierung des nystagmus und anderer augenbewegungen verm itteles des saitengalvanometers. *Deut Arch fur klin Med*, 140:79–90, 1922.

[SDL+05] Veronica Sundstedt, Kurt Debattista, Peter Longhurst, Alan Chalmers, and Tom Troscianko. Visual attention for efficient high-fidelity graphics. In *Proceedings of the 21st spring conference on Computer graphics*, pages 169–175. ACM, 2005.

[Sel13a] Elmedin Selmanovic. *Stereoscopic high dynamic range imaging.* PhD thesis, University of Warwick, 2013.

[Sel13b] Elmedin Selmanovic. *Stereoscopic high dynamic range imaging.* PhD thesis, University of Warwick, 2013.

[She92] James E Sheedy. Vision problems at video display terminals: A survey of optometrists. *Journal of the American Optometric Association*, 63(10):687–692, 1992.

[SHS+04] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. High dynamic range display systems. *ACM Transactions on Graphics (TOG)*, 23(3):760–768, 2004.

[SIM15] SIM2 Multimedia S.p.A. Sim2 high dynamic range display series. `http://hdr.sim2.it`, July 2015.

[Smi67] Bruce Smith. Geometrical shadowing of a random rough surface. *IEEE transactions on antennas and propagation*, 15(5):668–671, 1967.

[SSH05] James E Sheedy, Rob Smith, and John Hayes. Visual effects of the luminance surrounding a computer display. *Ergonomics*, 48(9):1114–1128, 2005.

[TKS06] Alejandro Troccoli, Sing Bing Kang, and Steve Seitz. Multi-view multi-exposure stereo. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 861–868. IEEE, 2006.

[Wan95] Brian A Wandell. *Foundations of vision.* Sinauer Associates, 1995.

[WFN+14] Junle Wang, Yuming Fang, Manish Narwaria, Weisi Lin, and Patrick Le Callet. Stereoscopic image retargeting based on 3d saliency detection. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 669–673. IEEE, 2014.

[WMLT07] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 195–206. Eurographics Association, 2007.

[WS13] Stefan Winkler and Ramanathan Subramanian. Overview of eye tracking datasets. In *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pages 212–217. IEEE, 2013.

[WW07] Andrea Weidlich and Alexander Wilkie. Arbitrarily layered micro-facet surfaces. In *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, pages 171–178. ACM, 2007.

[WWZ+09] Rui Wang, Rui Wang, Kun Zhou, Minghao Pan, and Hujun Bao. An efficient gpu-based approach for interactive global illumination. *ACM Transactions on Graphics (TOG)*, 28(3):91, 2009.

[Yar67] Alfred L Yarbus. Eye movements during perception of complex objects. In *Eye movements and vision*, pages 171–211. Springer, 1967.

[YJL17] Shun-nan Yang, Manho Jang, and Ju Liu. Neuro-behavioral effects of luminance level on visual performance and discomfort with high dynamic range displays. 2017.

[YPG01] Hector Yee, Sumanita Pattanaik, and Donald P Greenberg. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics (TOG)*, 20(1):39–65, 2001.

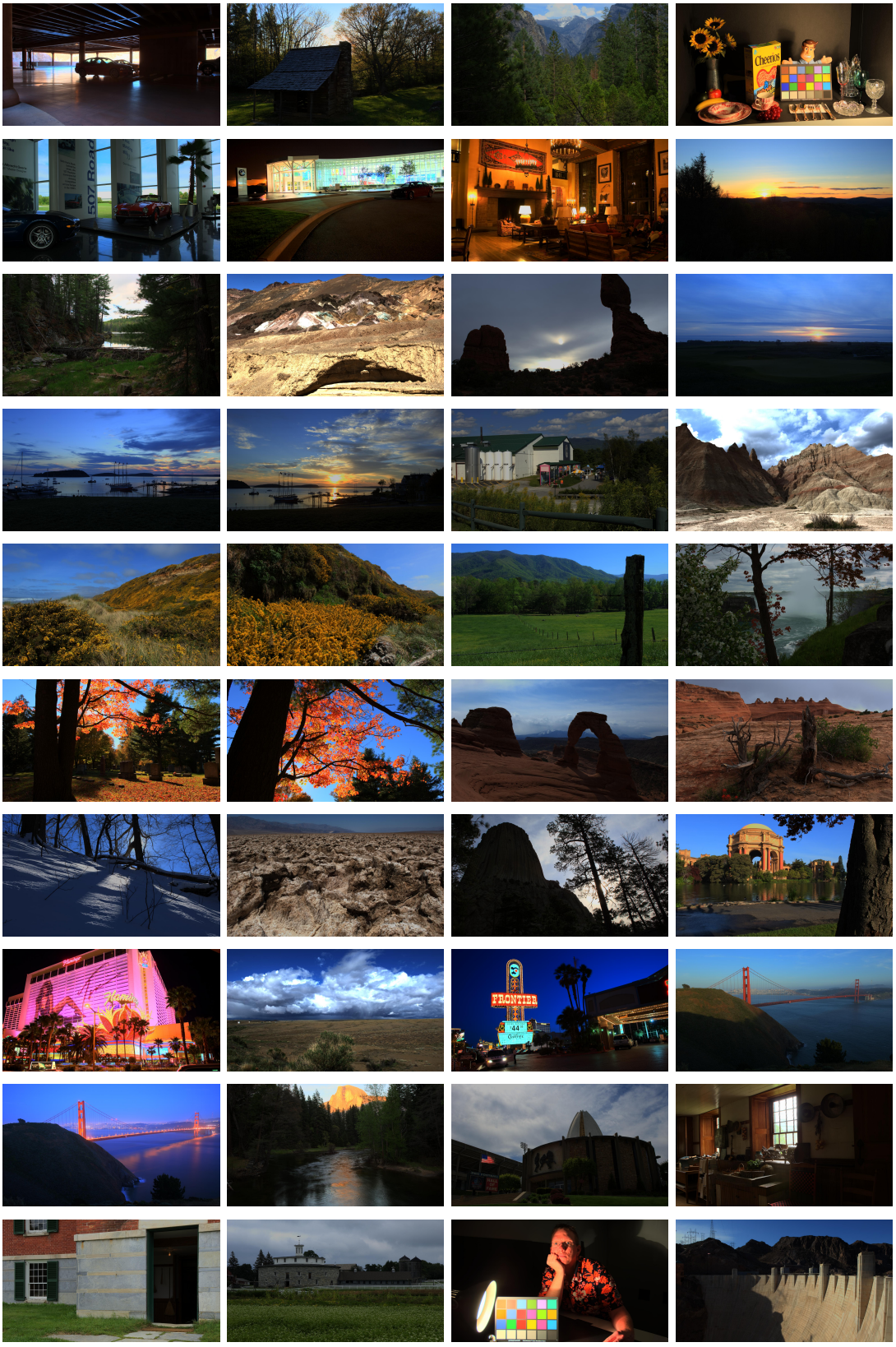
[YT89] PT Yeow and SP Taylor. Effects of short-term vdt usage on visual functions. *Optometry & Vision Science*, 66(7):459–466, 1989.

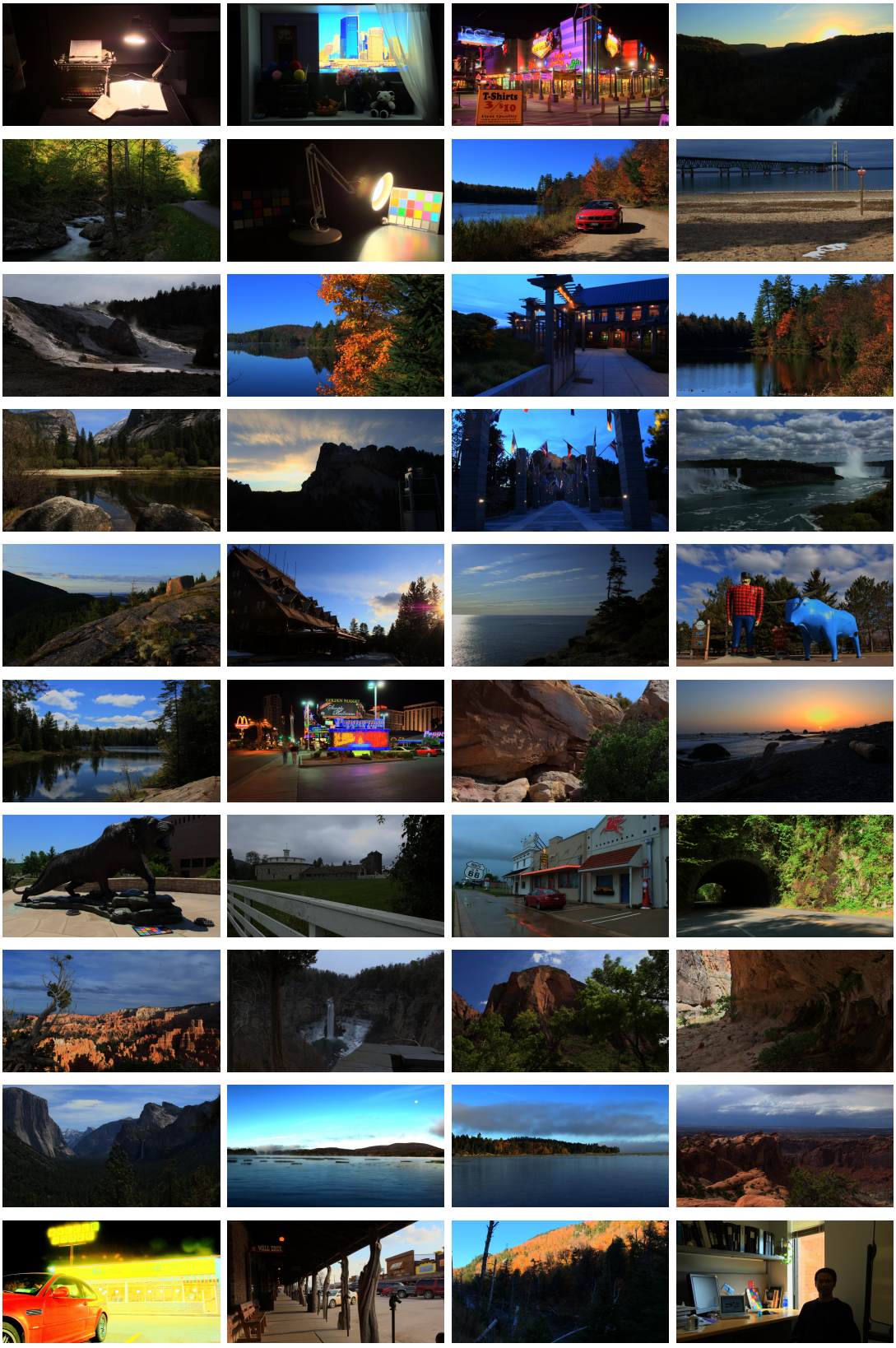
Appendices

Appendix A

Experiment One Images

These images were used in Chapter 6. They were all taken from the Fairchild Database [Fai07].

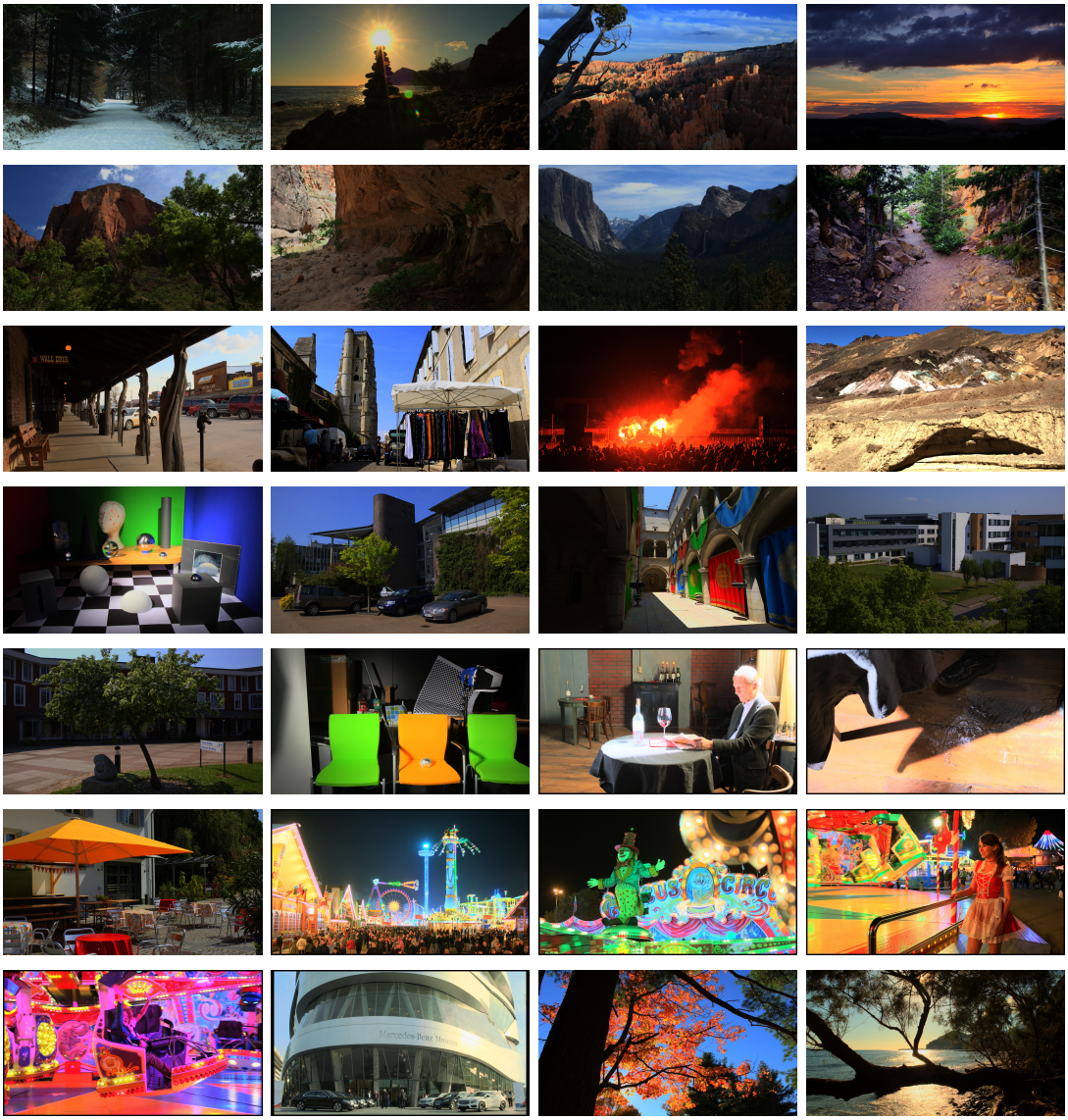

Cheerios
FRONTIER

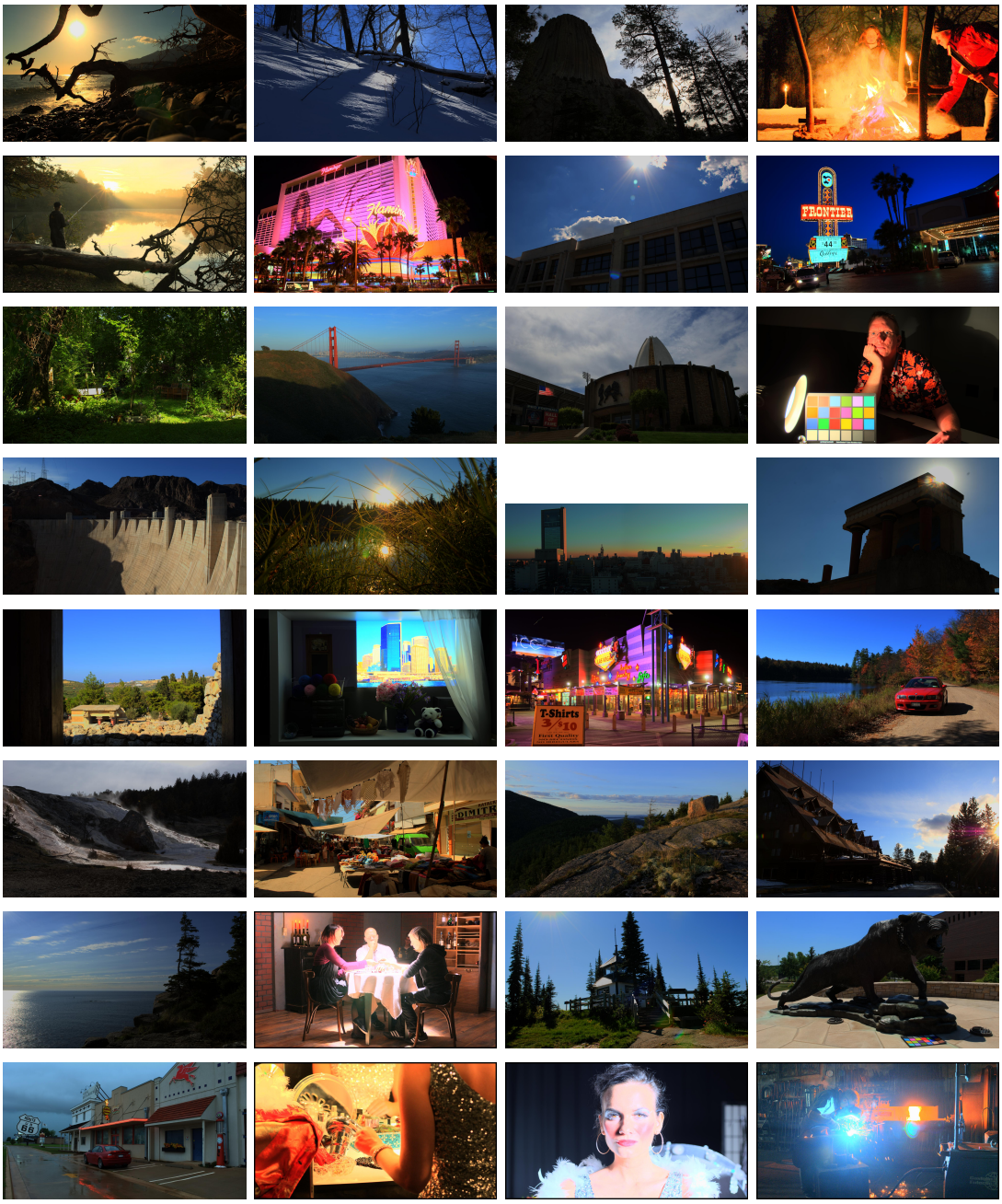

T-Shirts
3 for $10

Appendix B

Experiment Two Images

This appendix contain the images used in the experiment presented in Chapter 7. They are taken from the Fairchild database [Fai07], Technicolor's HDR dataset [LLF13], and the dataset of Froehlich et al. [FGE+14]




FRONTIER
T-Shirts
3 $10